

D6.6 Roadmap towards dynamic and automated building data collection

Third draft M36

Project acronym	BuiltHub
Full title	Dynamic EU building stock knowledge hub
GA no	957026
WP, Deliverable #	WP6, D6.6
Version	Draft M36
Date	30.09.2023
Dissemination Level	Public
Deliverable lead	Eurac
Author(s)	Ulrich Filippi Oberegger, Eurac
Reviewer(s)	Sympraxis
Keywords	Building stock, data collection, roadmap



This project has received funding from the EU's Horizon 2020 programme under grant agreement no 957026.

Disclaimer

The sole responsibility for the content of this publication lies with the authors. It does not necessarily reflect the opinion of the European Union. Neither the EASME nor the European Commission is responsible for any use that may be made of the information contained therein.

Table of contents

1. Introduction	7
2. The importance of a Building Stock Observatory	8
2.1. Context	8
2.2. Purposes and goals	14
2.3. Services	14
2.4. Features	15
2.5. Customers	16
2.6. Establishing a link between macro and microdata.....	18
3. The building data landscape of today.....	19
3.1. Introduction	19
3.2. Data inventory	19
3.3. Data quality and comparability	20
3.4. Data elaboration.....	22
3.5. In-depth study: EPC databases	22
3.5.1. Case study #1: England and Wales	24
3.5.2. Case study #2: The Netherlands	26
3.5.3. Case study #3: Denmark	27
3.5.4. Other countries	28
3.6. Data service providers	29
3.7. Existing European data strategies and roadmaps	32
4. Recommended measures for a self-sustained building data value chain.....	40
4.1. Political.....	40
4.2. Economic	42
4.3. Social	45
4.4. Technological	46
4.5. Legal	48
4.6. Environmental	48

5. Implementation and conclusion	49
5.1. Roadmap timeline in brief	49
5.2. Recommended steps to undertake	50
5.3. Conclusion	57
6. References	58

Abbreviations

Abbreviation	Description
AEC	Architecture, Engineering, Construction
AI	Artificial Intelligence
ANOVA	Analysis of Variance
API	Application Programming Interface
BIM	Building Information Modelling
BRP	Building Renovation Passport
BSO	Building Stock Observatory
CA	Concerted Action
CDM	Common Data Model
COP	Coefficient of Performance
CSV	Comma-Separated Values
DBL	Digital Building Logbook
DCAT-AP	Data Catalog Vocabulary Application Profile
EED	Energy Efficiency Directive
ESG	Environmental, Social, and Governance
EPBD	Energy Performance of Buildings Directive
EPC	Energy Performance Certificate
FAIR (data principle)	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
GHG	Greenhouse Gas
IEQ	Indoor Environmental Quality
INSPIRE	Infrastructure for Spatial Information in Europe
IP	Intellectual Property
ISO	International Organization for Standardization
KID	Key Indicators and Decisions
LAU	Local Administrative Unit
LTRS	Long-Term Renovation Strategy
ML	Machine Learning

NBRP	National Building Renovation Plan
NECP	National Energy and Climate Plan
NEEAP	National Energy Efficiency Action Plan
NREAP	National Renewable Energy Action Plan
NUTS	Nomenclature of Territorial Units for Statistics
OWL	Web Ontology Language
PDF	Portable Document Format
RED	Renewable Energy Directive
RDF	Resource Description Framework
ROI	Return on Investment
SEAP	Sustainable Energy Action Plan
SECAP	Sustainable Energy and Climate Action Plan
SFSB	Smart Financing for Smart Buildings
SRI	Smart Readiness Indicator
UML	Unified Modelling Language
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
WP	Work Package
XML	Extensible Markup Language

1. Introduction

The European Commission's efforts to design effective policies for buildings, as developed in the 2050 strategy and the European Green Deal, depend on detailed knowledge of building stock properties and behaviour. However, this can be challenging to manage, considering the lack of comprehensive and reliable data on the EU building stock.

In this report, the EU-funded BuiltHub project presents a roadmap for sustained data flow characterising the EU building stock, from data collection to exploitation, with the purpose to continuously inform and improve building-related policies and business through a community and its data hub. It seeks to positively disrupt policy and market decisions through a community-enhanced evidence base. The BuiltHub community representing the range of building stakeholders across Europe and its IT platform serving as data analytics and knowledge exchange hub shall change the way knowledge on the EU building stock is developed and shared.

Section 2 of this report states the challenge and importance of a Building Stock Observatory (BSO). Section 3 gives an overview of the building-related data collection, sharing, and valorisation framework and landscape of today by mapping key stakeholders and players as well as building data value chain processes. Section 4 presents BuiltHub's roadmap to support the transition of this landscape to its next generation. Measures are proposed to overcome existing barriers and create added value out of data by offering targeted services to stakeholders based on sound business models. In Section 5, implementation steps for the proposed measures are given and conclusions drawn.

This document is a draft version. There will be annual updates in parallel with the overall development of the project's results. The final version will be delivered at the end of the project scheduled for September 2024.

2. The importance of a Building Stock Observatory

2.1. Context

Many European initiatives, directives, and policies, among other measures, aim to transform the European building stock to combat climate change, support the economy, reduce energy poverty, and ultimately provide better life conditions to all Europeans. The following list mentions some of them.

- **Strategies and initiatives**
 - Several energy topics declared by the European Commission¹, such as “Energy Efficiency”, “Renewable Energy”, and “Oil, gas and coal” (efficient and responsible use of fossil fuels), touch the building sector.
 - Since 2015, the “Energy Union strategy”² aims to ensure energy security, a barrier-free internal energy market, energy efficiency, decarbonisation based on renewable energy, and support of low-carbon and clean energy technologies.
 - The Clean Energy for all Europeans package³ published in 2016 is an important step towards implementing the energy union strategy focused on:
 - Building energy performance through the amending Energy Performance of Buildings Directive (EPBD) of 2018⁴ and its revision, which is in progress.
 - Renewable energy through the Renewable Energy Directive (RED) with a revision proposed in 2021⁵ and a provisional agreement “to raise 2030 target to at least 42.5%, aiming for 45%” reached on 30 March 2023⁶.
 - Energy efficiency through the Energy Efficiency Directive (EED) amended in 2018⁷.
 - Governance, requiring for instance Member States to submit integrated National Energy and Climate Plans (NECPs) at regular time intervals.

¹ https://ec.europa.eu/energy/topics_en. Accessed 19 October 2021

² <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2015:80:FIN>. Accessed 19 October 2021

³ <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1481278671064&uri=CELEX:52016DC0860>. Accessed 19 October 2021

⁴ https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2018.156.01.0075.01.ENG. Accessed 19 October 2021

⁵ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0557>. Accessed 19 October 2021

⁶ https://energy.ec.europa.eu/topics/renewable-energy/renewable-energy-directive-targets-and-rules/renewable-energy-directive_en#revision-of-the-directive. Accessed 6 September 2023

⁷ https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2018.328.01.0210.01.ENG. Accessed 19 October 2021

- “Electricity Market Design”, with a reform proposed on 14 March 2023⁸ towards more flexibility, open markets, and an increased share of renewables.
- The Smart Financing for Smart Buildings (SFSB) initiative launched as part of the Clean Energy for all Europeans package to unlock public and private funds for energy efficiency and renewable energy in buildings.
- The European Green Deal aims at drastic decarbonisation, with milestones in 2030 and 2050.
 - The European green bond standard⁹ (EUGBS) is a voluntary standard for adoption by issuers and investors of green bonds. The EUGBS is supposed to provide enhanced alignment with the EU legislative framework, transparency, reliability, checking, and supervision.
 - The 2030 Climate Target Plan sets out targets for 2030 on a pathway in line with the European Green Deal.
 - The Fit for 55 package¹⁰ presents revisions and initiatives linked to the European Green Deal climate actions and in particular the climate target plan's 55% net reduction of greenhouse gas emissions (compared to 1990) by 2030.
 - The Renovation Wave is a key initiative in the European Green Deal to boost annual building energy renovation rates.
 - The EU Strategy for Energy System Integration as part of the European Green Deal is devoted to the energy system cross-sector coupling (buildings, transport, industry).
 - REPowerEU¹¹ includes investing in renewable energy, which includes using building roofs and façades to their full potential for photovoltaics.
- The COP26 of the UNFCCC¹² has seen the EU in the role as admonisher of low-transparency practices in green finance including ESG in green bonds.
- **Directives and regulations**
 - The Energy Performance of Buildings Directive (EPBD) has gone through several iterations over the years and is a cornerstone in the transition to a highly energy efficient and decarbonised building stock. In its Article 2a, it requires each Member State to establish a Long-Term Renovation Strategy (LTRS) for renovating the national building stock towards high energy efficiency and decarbonisation. Article 6 states the provisions for new buildings and specifically the nearly zero energy and high share of renewables targets, ensuring that high energy efficiency systems such as decentralised supply from

⁸ https://energy.ec.europa.eu/topics/markets-and-consumers/market-legislation/electricity-market-design_en. Accessed 6 September 2023

⁹ https://ec.europa.eu/info/business-economy-euro/banking-and-finance/sustainable-finance/european-green-bond-standard_en. Accessed 8 November 2021

¹⁰ <https://www.europarl.europa.eu/legislative-train/theme-a-european-green-deal/package-fit-for-55>. Accessed 19 October 2021

¹¹ https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal/repowereu-affordable-secure-and-sustainable-energy-europe_en. Accessed 7 September 2023

¹² <https://unfccc.int/conference/glasgow-climate-change-conference-october-november-2021>. Accessed 11 September 2023

renewable sources (e.g. PV), district heating, heat pumps etc. are considered. Concerning existing buildings, Article 4 requires Member States to set minimum energy performance requirements for buildings and components within buildings following the cost-optimality methodology. Article 8 addresses proper optimisation of technical building systems including smartness, and electromobility. Member States must fulfil additional obligations on reporting, control and monitoring of implementation, which are subject to penalties in case of non-compliance. Contributing measures such as financing, overcoming of market barriers, Energy Performance Certificates (EPCs)¹³, and building systems inspection are given as well. The EPBD is currently under review.

- The Energy Efficiency Directive (EED) sets energy efficiency and energy consumption reduction targets to be reached by EU countries. The directive requires each Member State to set national energy efficiency targets, based on either primary or final energy consumption, primary or final energy savings, or energy intensity. Member States also must show how these targets have been calculated. Furthermore, the EED adds a specific requirement for the public sector to lead by example through achieving a 3% renovation rate for publicly owned buildings and significantly reduce emissions.
 - The Renewable Energy Directive (RED) aims to develop renewable energy technologies and deployment across all sectors in the EU by converting into law approaches such as the Energy System Integration strategy.
 - The INSPIRE (Infrastructure for Spatial Information in Europe) Directive is central in achieving BuiltHub's objectives in that it aims to bring together spatial data across EU Member States addressing 34 *spatial data themes* for environmental purposes including buildings. A toolkit ranging from data specifications to good practices and trainings helps with the implementation.
 - The Commission Implementing Regulation (EU) 2020/2156 of 14 October 2020 expands on the concept of Smart Readiness Indicator (SRI)¹⁴ introduced in the 2018 amendment of the EPBD. The SRI's purpose is to rate the smart readiness of buildings, i.e., the capability of buildings or building systems to adapt their operation based on the behaviour and needs of occupants, the grid, and energy efficiency. As set out in the Regulation, professionals (e.g., energy performance certifiers) may be trained and accredited for issuing SRI certificates. These may be integrated with the EPC and with inspection schemes. Further, monitoring, promotion, and testing of the SRI are regulated.
- **Plans**
 - As set out in the EPBD 2018 Article 2a, each EU country had to develop a Long-Term Renovation Strategy (LTRS) for the national building stock, with a perspective of at least 30 years. The LTRS had to include:
 - An overview of the building stock. This typically includes an analysis of the distributions in terms of building type (residential, commercial, etc.),

¹³ https://ec.europa.eu/energy/topics/energy-efficiency/energy-efficient-buildings/certificates-and-inspections_en. Accessed 8 November 2021

¹⁴ <https://smartreadinessindicator.eu/>. Accessed 8 November 2021

age, size, tenure, occupation, and a clustering of the stock into homogeneous groups to establish the most appropriate renovation measures for each group. Clustering parameters may include climate, construction period, building geometry or size, building systems.

- Measures for cost-effective deep renovation of buildings. Cost-effectiveness implies that the worst performing buildings are addressed in a timely fashion. The development of these measures requires reliable data and knowledge about the building envelope, technical systems, and energy performance.
- Measures targeting energy poverty, market and other barriers (e.g., the split incentive problem), and public buildings.
- An overview of national initiatives regarding technologies, training, and education in related sectors.

Note: in the revision in progress of the EPBD, the LTRS are being replaced by National Building Renovation Plans (NBRP).¹⁵ An explanation and further details on NBRPs are given below.

- The National Energy and Climate Plans (NECPs) are the EU countries' 10-year national energy and climate plans for 2021-2030. The LTRS is part of the NECP.
 - In the National Renewable Energy Action Plans (NREAPs) EU countries set out their plans to fulfil at least 20% of the total energy needs with renewables by 2020.
 - The National Energy Efficiency Action Plans (NEEAPs) are drawn up by EU countries about every three years. The progress achieved towards their national energy efficiency targets is to be reported on an annual basis.
 - SEAPs (Sustainable Energy Action Plans) are plans drawn up by signatories of the Covenant of Mayors. These are administrations of towns, cities, and regions who committed to reducing their CO₂ emissions beyond the 20% target for 2020.
 - Inspired by the SEAPs, the SECAPs (Sustainable Energy and Climate Action Plan) extend and update the former plans to a comprehensive view combining energy and climate targets for 2030.
 - The European Strategic Energy Technology Plan (SET Plan) is devoted to the development of low-carbon technologies and has created 10 key actions revolving around renewables, energy systems, energy efficiency, transport, carbon capture and nuclear safety.
- **Frameworks**
 - Level(s)¹⁶ allows assessing buildings throughout their life cycle based on six macro-objectives tracked through 16 indicators on greenhouse gas emissions, materials use, water use, indoor health and comfort, adaptation and resilience to climate change, and optimised life cycle cost and value. By adopting Level(s),

¹⁵ https://www.europarl.europa.eu/doceo/document/TA-9-2023-0068_EN.html. Accessed 07/06/2023

¹⁶ https://ec.europa.eu/environment/levels_en. Accessed 9 November 2021

developers, policy makers and designers can use standardised indicators to evaluate and compare buildings throughout their life cycle.

- **Actions**

- The EU established several Concerted Actions (CA) to support the transposition and implementation of directives including the EED, EPBD and RED (see above). The central link with BuiltHub is given by the CA-EPBD (in close contact with CA-RES and CA-EED) whose participants involve representatives of national ministries or their affiliated institutions in charge of preparing the framework for the implementation of the EPBD in each Member State, including the UK and Norway. Core Teams (CT) and Cross-Cutting Teams (CCT) focus on specific topics.

Central to the success of these efforts is the need for reliable, transparent, accessible, and accurate data and metadata. Focusing on buildings, the “EU Building Stock Observatory (BSO) was established in 2016 as part of the Clean energy for all Europeans package and aims to provide a better understanding of the energy performance of the building sector through reliable, consistent and comparable data.” The declared goal of the BSO is to “support monitoring the implementation of different measures and contribute to future policy making. The data published in the BSO can therefore be very useful to policymakers, investors, stakeholders, local and national authorities, and researchers.”¹⁷

The BSO has gone through several iterations since its inception, which showed the importance and the relevance of the endeavour but also the numerous challenges and barriers to overcome. An analysis of the second iteration of the BSO is found in BuiltHub deliverable D3.1 “Inventory structure and main feature and datasets”¹⁸. In that deliverable, data gaps have been identified as one of the most pressing limitations of the BSO. It has been concluded that developing a roadmap for enhanced, more comprehensive, and continuous data provision is one of the most important actions needed to improve the BSO service. The deliverable also provides details concerning specific data gaps and data availability, pointing out spatial/geographic gaps (i.e., data for several countries is missing) as well as temporal gaps (gaps in the time series), and indicators for which only limited data is provided, such as smart metering, certification, financing, and the energy market.

Further points mentioned in D3.1 include the necessity to integrate new indicators considering that energy system and society are rapidly evolving and have a strong impact on our lives in the buildings. Such thematic areas include renewables, flexibility and smart readiness, e-mobility, and IEQ (Indoor Environmental Quality), among others. Also, the indicator descriptions, infographics, and ways to download and compare data can be improved. As identified in D3.1, an important barrier to overcome is insufficient metadata. Comprehensive metadata would allow users to judge the suitability and exploitability of data for their purposes.

¹⁷ https://ec.europa.eu/energy/topics/energy-efficiency/energy-efficient-buildings/eu-bso_en. Accessed 05/07/2021

¹⁸ https://builthub.eu/fileadmin/user_upload/DELIVERABLE_D3.1.pdf. Accessed 04/10/2022

To overcome these barriers, the EPBD continues being revised. On 14 March 2023, the European Parliament has adopted amendments on the proposal for recasting the directive.¹⁹ In Article 19 of this adopted text on databases for energy performance of buildings, it is stated that:

“1. Each Member State shall set up a national database for energy performance of buildings which allows data to be gathered on the energy performance of *individual* buildings and on the overall energy performance of the national building stock.”

“4. At least once per year, Member States shall ensure the transfer of the information in the national database to the Building Stock Observatory.”

Concerning renovation, a core goal in the decarbonisation of the European building stock, Article 3 on National Building Renovation Plans (NBRP) states:

“1. Each Member State shall establish a national building renovation plan to ensure the renovation of the national stock of residential and non-residential buildings, both public and private, into a highly energy efficient and decarbonised building stock by 2050, with the objective to transform existing buildings into zero-emission buildings.”

The text explicitly mentions that these “national building renovation plans [...] replace the long-term renovation strategies”.

Further, in Article 3, a direct link is established with Regulation (EU) 2018/1999²⁰ on planning and reporting obligations across energy and climate policy areas.

How Member States are supposed to report progress in these policy areas is laid out in the Commission Implementing Regulation (EU) 2022/2299 of 15 November 2022²¹. Annex IV, Table 2 on “Milestones and progress indicators of the long-term strategy for the renovation of the national stock of residential and non-residential buildings – building stock” asks for the following “mandatory if available” (M_{iav}) data, where M_{iav} “means a category of information that Member States have to submit only if such information is available to them at the time of the submission of the biennial progress report.”

- Number of buildings renovated;
- Total floor area renovated;
- Primary energy use of buildings;
- Final energy use of buildings;
- Direct GHG (greenhouse gas) emissions in buildings;
- Total GHG emissions in buildings

This data is to be provided for the years 2020, X-3, and X-2, where X denotes the reporting year, and separately for residential, non-residential, and public buildings.

¹⁹ https://www.europarl.europa.eu/doceo/document/TA-9-2023-0068_EN.html. Accessed 07/06/2023

²⁰ <http://data.europa.eu/eli/reg/2018/1999/2023-05-16>. Accessed 07/06/2023

²¹ http://data.europa.eu/eli/reg_impl/2022/2299/oj. Accessed 07/06/2023

Table 3 goes into more detail on the reporting of the energy renovation depths achieved. Three depths are defined: “light” (3 to 30% savings), “medium” (30 to 60% savings), and “deep” (a renovation which transforms a building or building unit (a) before 1 January 2030, into a nearly zero-energy building (b) as of 1 January 2030, into a zero-emission building).

2.2. Purposes and goals

“If you can’t measure it, you can’t improve it.”²² The basis for being able to improve the European building stock is to know its features in detail and how it evolves. Therefore, the first and foremost purpose of a BSO is to provide reliable and relevant information about the building stocks of the countries and how they are changing year over year. However, many more purposes and goals can be associated with a future BSO. These must adequately address the needs and requirements of all stakeholders in Europe. Consequently, the following purposes have been identified in a series of collaboration and engagement efforts with stakeholders, see also BuiltHub deliverables D2.1 “Stakeholder mapping” (only available on the EC Participant Portal) and D2.2 “Report on relationships of stakeholder needs and requirements”²³:

- Inform about the building stocks in the different countries
- Provide raw data for further research and analysis
- Support the development of renovation, decarbonisation, and investment plans
- Monitor the implementation and effectiveness of different measures
- Contribute to policy and decision making
- Accelerate analyses by providing ready-to-use analytics, knowledge, and insight through, e.g., dataset comparisons, key performance indicators, projections, and scenarios
- Provide benchmarking values for the construction and related sectors (e.g., real estate, e-mobility, automation, renewables), market developments (e.g., energy prices, labour cost, construction materials cost), industrial products, and technology learning curves (e.g., batteries)

2.3. Services

To be able to effectively fulfil the identified purposes, the next-generation BSO should strive to offer the following basic services:

- An open, central, easy, one-stop access to building stock data relevant to the stakeholder’s mission
- Comprehensive, transparent data descriptions and metadata
- User-friendly possibilities to export historic as well as up-to-date data (with few operations, retrieving exactly what is needed and in a suitable format)

²² The original sentence attributed to the British scientist Lord Kelvin (1824-1907) expresses this idea in a more convoluted way.

²³ https://builthub.eu/fileadmin/user_upload/BuiltHub_deliverable_D2.2_Final.pdf. Accessed 04/10/2022

- User-friendly possibilities to visualise data through different kinds of charts
- Aggregation operations and descriptive statistics (totals, means, medians, percentiles, histograms, box plots, etc.)
- Comparisons of datasets to assess data consistency
- Relevant key performance indicators
- Spatial data and indicators representation at the respective NUTS or LAU levels

The above BSO services should connect with other existing services or services under construction. For instance, an EU framework for Digital Building Logbook (DBL) gateways (R2M Solution, VITO, BPIE, 2020) and related provisions are being developed. In this context, “gateway” refers to the possibility that the DBL itself collates data and information from various data sources and is kept up to date whenever data is being updated at the source. Data sources of interest may include EPC databases, registers, cadastres, BIM data, financial data, Level(s) indicators, bills of materials, etc.

More advanced services and features may be considered for future implementation:

- Projections (trends) for a variety of projection methods fitted to historic data or additional available data, e.g. technology learning curves
- Automatic connections with other databases
- Automatic periodic data updates
- Programmatic access to raw data and analytics
- Calculation of scenarios based on customisable models
- Benchmarking, i.e., indicators for a user-selected data subset are compared with benchmark values for these indicators

In BuiltHub, a web-based platform was created to test and demonstrate such services to inform further development and refinement of the EU BSO. The platform is accessible through the project’s website²⁴. Feedback loops with external stakeholders ensure that the platform’s services address their requirements properly.

2.4. Features

To best fulfil the purposes and reach the goals stated in Section 2.2 as well as deliver the services in Section 2.3, a series of characteristics are recommended for the next-generation BSO:

- Accessibility of relevant data to all stakeholders, anytime
- Scalability – increasing the amount of data, analytics, and stakeholders accessing the platform should not deteriorate the platform’s performance
- Modularity, flexibility, and extensibility for platform developers thanks to several layers of abstraction
- Well-documented APIs to establish connections with other data platforms or data analytics services

²⁴ <https://builthub.eu/about/the-platform>. Accessed 12 September 2023

- Possibilities for the user community to give feedback about bugs and features and to discuss, in a kind of forum, thematic improvements
- Standardised data models, descriptions, and metadata – it should be clear (unambiguous) what the data represents
- User-friendly frontends, data browsing and export experience
- Possibilities to easily provide data – a data provider must have appropriate means to be able to periodically upload or make accessible its data to the BuiltHub platform
- Adequate licencing, allowing ease of data re-use for the platform user within an economically viable business model for the data provider
- Pre-calculated indicators, plans, and scenarios, or calculations on-the-fly providing analytics, intelligence, and insight addressing the needs and requirements of stakeholders
- Flexible spatial representation of data at different levels of granularity
- Sufficiently realistic, but especially transparent (well-documented) calculation models and underlying assumptions
- Open-source code and appropriate contracting and licence models for the BSO itself, or parts of it

The web-based platform created in BuiltHub (see also Section 2.3) aims to implement these principles and features to serve as test bench and provide guidance in the development of future generations of the BSO.

2.5. Customers

Customers have been grouped into three stakeholder types, namely **Lead-users**, **End-users**, and **Multipliers**, as reported in deliverable D2.2 “Report on relationships of stakeholder needs and requirements”, Figure 1. This subdivision is loosely based on the main role of the customer, but considerable overlaps are expected. Each group consists of subgroups, which still represent extensive sectors:

- Lead-users (data providers)
 - Research
 - Architecture, Engineering, Construction (AEC)
 - Energy services
 - Property and facility management
 - Statistics institutes
- End-users (platform users)
 - Government
 - AEC
 - Regulation and standardisation
 - Housing organisations
 - Individual owners, citizens
- Multipliers (promoters)
 - Government networks, stakeholder associations, and EU agencies
 - Property owner associations
 - Civil society and media

In summary, in addition to **private or public stakeholders**, the **European Commission** and the **EU Member States** are among the clients of the BSO. The European Commission can use information about the building stock transition in the different countries to assess the effectiveness of its directives, regulations, and policies. It can also monitor how well Member States have implemented energy efficiency measures, e.g., whether and how much the energy performance of their building stock increased. Key performance indicators such as building shell performance can give an indication as to why the energy performance increased or decreased. Combining with investment data and appropriate indicators can give further insight whether money was spent effectively.

EU Member States can leverage the data in the BSO by basing their analyses on or integrating it in their LTRS (Long-Term Renovation Strategies) and NECP (National and Energy Climate Plans). Commonly, EU Member States may delegate the elaboration of such plans and strategies to official entities in their country, such as **energy agencies**. These in turn will base the development of the data, models, documents, and reports to be delivered on reliable datasets coming from, e.g., **national statistics institutes**. Oftentimes they ask **research institutions, market research bureaus**, and other experts for support and advice in specific areas. Therefore, the BSO could provide these stakeholders with reliable, standardised, comparable reference data. This would enable them to shift their attention from the data collection, which today might very well take most of the available resources, to the analysis and elaboration of strategies and recommendations for improvement.

By adding data for different spatial scaling levels (not only the national or NUTS0 level), the BSO can attract additional customers, such as **building portfolio managers**, both public and private. The data and services of the BSO might be used by them and by **public authorities** to benchmark the performance of their own policies and measures against other countries, regions, cities, or in general building clusters, and to demonstrate and showcase effective measures others can study and learn from. They might also want to use results from the BSO to motivate investment decisions.

On the other hand, all the above-mentioned clients can act as **data providers**. It is therefore important that the BSO allows for easy integration of new data while keeping highest data quality and reliability standards.

Major hurdles to overcome in the data provision process are appropriate **licensing, ownership protection, and IP (Intellectual Property) protection**. Concerning the BuiltHub platform, these aspects are outlined in the confidential BuiltHub deliverable D1.3 “Data Management Plan” and investigated in-depth in several other confidential project deliverables. Two main goals must be reached:

1. **Protect** the (data and intellectual) property of all parties involved
2. **Create incentives** for all parties to proactively contribute to the data value chain through adequate business approaches

Data and IP protection is rendered complex since any step in the data value chain or DIKW (Data, Information, Knowledge, Wisdom) Pyramid may involve different owners, each contributing with their own property. For instance, metadata, aggregated data, analytics, ML algorithms, and other results can have different ownerships.

In this report, we will give a summary of the conclusions reached and shareable recommendations derived from the above-mentioned confidential deliverables.

By integrating big data and microdata into the BSO, numerous additional clients may find interest and regularly use the BSO, see Section 2.6.

2.6. Establishing a link between macro and microdata

The BSO may potentially serve the purpose to link macro and microdata together. With *macro data* we refer to highly aggregated data, either spatially or temporally. Because of their more generic nature, they may be available for many spatial regions (e.g., all EU countries) or time periods (e.g., years). *Microdata* on the other hand have a high spatial or temporal resolution (or both) but may be available only for specific case studies or few spatial regions or time frames. As of today, the BSO is hosting only macro data, i.e., annual data at EU and mostly country level, with single regions²⁵ featured as exception. However, a high amount of microdata, frequently associated with big data, is available but arguably largely untapped that could complement the macro data in the BSO and make the BSO attractive for many additional stakeholders. One opportunity in leveraging isolated sets of microdata is bringing them together for higher statistical representativeness. Further, microdata typically contains detailed, valuable information not available in macro data. However, data pertaining to a small sample may not generalise to the entire building stock's properties.

Making such isolated, detailed datasets accessible through the BSO allows gradually bringing them together until statistical representativeness is reached. It then becomes possible for stakeholders to generalise the information contained in these datasets to a more extensive spatial region or time frame with high confidence, thus allowing a bottom-up comparison with top-down macro data. Such a comparison would serve as a further validation of the macro data. At the same time, important variations and differences in the various micro datasets are kept and can be analysed. Microdata further allows flexible re-use for different applications and contexts.

Even if generalisation or extrapolation is not possible, microdata can be aggregated to serve as benchmark or reference intelligence for stakeholders and other datasets becoming available later.

By integrating big data and microdata into the BSO or by providing access to such data from the BSO, many additional stakeholders, such as **utilities, aggregators, building (facility, energy) managers, building materials, component, and equipment manufacturers, designers, building industry companies, real estate companies, building portfolio managers, and local authorities**, may want to become regular users of the BSO.

Data layers with intermediate levels of detail are highly recommended to efficiently bridge the gap between microdata and macro data. If we consider microdata as close-to-raw but clean and high-quality data with no substantial level of aggregation, a main additional layer towards higher aggregation consists of datasets storing KPIs at single building and annual level. Digital

²⁵ As of 8 July 2021: England, Flanders, Northern Ireland, Scotland, Wallonie, Wales

Building Logbooks (DBLs) could constitute such a type of dataset. Through appropriate aggregation (i.e., the calculation of totals, averages, and the like), these KPIs can be calculated at building cluster level, where a cluster consists of a sufficiently high number of buildings (e.g., 100) to allow compliance with **privacy** requirements, the General Data Protection Regulation (GDPR), and data ownership protection. This data at building cluster level forms the next main layer and can be shared and reused more easily. From this point onward, a generalisation to different LAU or NUTS levels through several further aggregation steps is straightforward and ultimately allows a comparison with macro data.

3. The building data landscape of today

3.1. Introduction

Deliverable D3.1 gives an overview of the indicators and data present in the BSO. It assigns a priority level to each indicator (low, medium, or high) based on its importance in describing or monitoring the building stock and on the feasibility to obtain data for its calculation. The latter criterion is important to ensure that an indicator can be computed at reasonable cost (depending on the benefits it can bring).

It also emerges that there are several data gaps in the BSO for certain thematic areas such as technical building systems, smart metering, EPCs (Energy Performance Certificates), financing, and energy market. This is arguably due to the difficulty to find and access data of sufficient quality for these areas and not because data is not available at all.

Where data is not available for specific geographic areas or past years, a variety of techniques, for example Machine Learning (ML), can be used to estimate missing data, which allows delivering highly valuable services to stakeholders. For instance, municipalities or larger regions as well as public or private real estate managers need support in understanding how their building stock evolves over time. Typically, they have detailed data about certain characteristics of the building stock, for example from their registers, but they miss other data that may be available in the database for similar contexts. ML techniques to estimate missing data are explored in the confidential D4.4 “Description of a workflow how to implement and transform data”.

Another issue is that a large amount of available data is highly fragmented, not openly accessible, and in a format making it difficult to leverage.

The following sections report in more detail how BuiltHub analysed and mapped out the building data landscape, from creating the data inventory (Section 3.2) to data quality assurance (Section 3.3) and data elaboration (Section 3.4). Section 3.5 provides an in-depth study of EPC databases.

3.2. Data inventory

The data inventory provides a complete list of data sources and datasets by leveraging best practices in information collection, such as scraping of national statistical data. Once a first

version of the data inventory is ready, a major challenge to further improve is to identify data gaps and to provide recommendations on how to close them.

One important aspect of the data inventory is to ensure that the data can be understood and interpreted correctly by any user. This requires a compilation of clear data description, annotation, contextual information, and documentation.

An essential step in finding all data sources is the proper definition of a research strategy. In general, an inventory of data is comprised of an overview of existing values, classified according to specific key themes and subthemes, a data documentation, and a metadata file on datasets. Creating a comprehensive data documentation involves collecting a large amount of information. Essential information includes:

- Features of the dataset, such as name, labels, data types, data descriptions, and issuing organization.
- Characteristics of the variables like provider, date of creation, methodology/definition, data manipulations, classification, missing values, reason for missing values, and descriptions and meanings of data.
- Metadata, which is a subset of core data documentation that should provide standardised structured information explaining the purpose, origin, time references, geographic location, creator, access conditions, and terms of use of the dataset.

The BuiltHub data inventory of 30 datasets considered important for the BSO is reported in the public Deliverable 3.1 “Inventory structure and main feature and datasets”.

The datasets include results from European projects (Tabula, Episcopo, Zebra2020, Entranze, Odyssee-Mure, HotMaps, CommONEnergy, etc.) as well as databases (JRC IDEES, Eurostat, Census, etc.) that should be consulted to collect data and perform a comparison against. The detailed metadata, including links to the datasets, is reported in the Deliverable 3.1.

3.3. Data quality and comparability

Data quality, completeness, accuracy, handling, and reliability are very important aspects to which all necessary attention needs to be dedicated. Even though most data providers utilize standardised data formats, this does not necessarily mean that data is reliable and comparable. Differences in definitions of the values and survey methodologies can contribute to an irrational and incomplete data comparability and consequently misinterpretation. In this respect, introducing measures to improve data comparability is one of the most important aspects of the whole process of data elaboration. It is a complex process of adjusting differences and inconsistencies among different measures, methods, schedules, and specifications to make them uniform and comparable.

Another important issue regarding data quality is the fact that definitions of quantities such as floor area, energy performance, etc. can vary across countries, or even across years for the same country. Providing clear definitions is a key aspect in the process of data comparability. To make variables (e.g., X, Y) coming from different sources and years comparable, an understanding of the meaning of each variable definition is required. Once a deep understanding is reached, different definitions can be compared.

The main problems with definitions concern the fact that because most of the data is collected from national statistics or national bodies, each national state defines parameters in particular ways, due to specific territorial and socio-economic conditions. Definitions, meanings, and terminologies involving certain issues, such as how floor area is accounted for (e.g., differences in gross, net, heated, useful floor area), what kind of energy performance is considered (e.g., primary, final, delivered), how labels in EPCs are assigned, etc., need to be checked and addressed before data may get aggregated. It is necessary to check and confront the meaning of definitions across different times and countries, review the methodology of data collection, sources utilized and metadata changes.

Deliverable D3.2 “Methodology on quality assurance” reports on the quality assurance procedures recommended by the BuiltHub consortium. Specifically, the deliverable shows how to support final users in their decision on which data to use for an indicator if there are several datasets available for that indicator. As shown in Figure 1, the first step of the quality control consists of thorough metadata collection allowing users to check where data comes from, thereby increasing their confidence. The second step is split in 3 levels. Level 1 is the most basic and only guarantees the provision of at least one dataset containing data for indicators requested by stakeholders. Level 2 incorporates a consistency analysis between two datasets, carried out through a linear regression and evaluated using standard statistical parameters (correlation coefficient, R^2 or coefficient of determination, adjusted R^2 , standard error, sample size), an ANOVA (analysis of variance), and a t-statistic readily available in statistics software but also widespread spreadsheet applications such as MS Excel. As suggested rule of thumb,

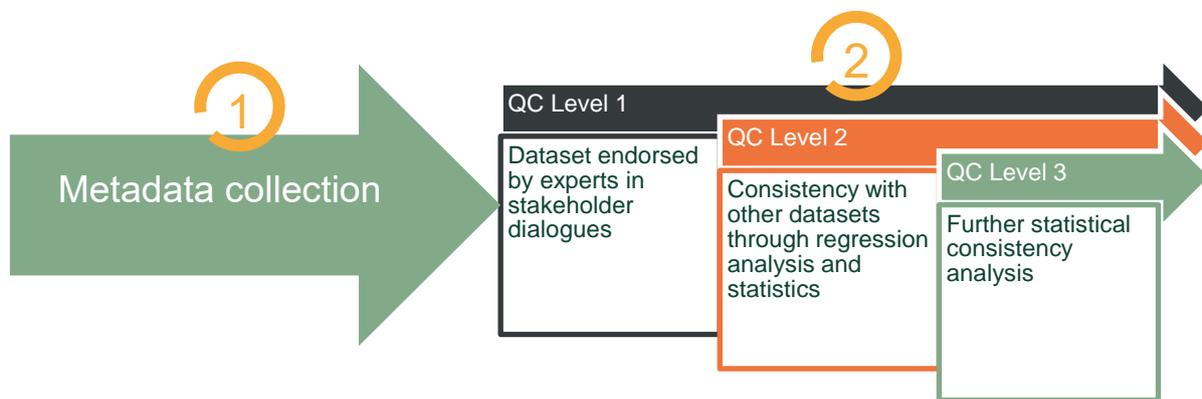


Figure 1: Overview of methodology on quality assurance

the two datasets are deemed consistent with each other if the p-value of the t-statistic is below 5%. Clearly, this threshold can be modified to meet desired accuracy requirements. Level 3 adds some descriptive statistics to the consistency analysis (minimum, maximum, standard deviation, quartiles). Finally, the deliverable gives several examples in which the quality control is applied to the BuiltHub datasets.

3.4. Data elaboration

Due to the specific degree of implementation, each target country has a different level of data availability. The extensive and comprehensive data collection and elaboration for each target country enables a comparative gap analysis.

Filling in the data gaps implies not only extrapolating and assembling data from large data tools available online (e.g., ENTRANZE²⁶), but also analysing data source-by-source from single scientific literature sources – such as journal papers, conference proceedings and project deliverables. Only through such an in-depth approach, gaps of data per energy type and country can be remedied.

To fill remaining gaps, data extrapolation techniques can be used using other gathered information, such as climate conditions, buildings construction typologies and age, GDP, and economy sectors.

If proper descriptions are missing concerning how the collected data have been obtained, those data need to be excluded from the database.

One of the major challenges in developing an inventory of data for different sectors is to provide a complete list of all already existing data. In general, the advantage of using data coming from EU wide data providers is that the data is available for a large territory (e.g., Odyssee Database²⁷). However, a drawback is that these EU-wide data providers tend to have incomplete databases. For this reason, it is recommended to collect additional/complementary data from national statistics and to create a list of national contact points that can offer support in data collection.

This approach allows better coverage of the available data in each country. The great effort directed towards data collection between countries and facing the problems of administrative and definition changes are the most challenging aspects of data inventories.

Information needs to be collected for recent years. The access to up-to-date and accurate data on the building stock and its energy is of fundamental importance for the assessment.

3.5. In-depth study: EPC databases

A core idea behind Energy Performance Certificates (EPCs) is to provide information to individuals planning to renovate, purchase or rent a building. An EPC typically includes an energy performance rating and recommendations for cost-effective improvements. EPCs can be a useful resource for policy development if they meet quality and reliability requirements.

However, the potential of EPCs strongly depends on their implementation and market coverage. The X-tendo project^{28,29} works on a toolbox for next generation EPCs and published

²⁶ <https://www.entranze.eu/tools/interactive-data-tool>. Accessed 29 March 2022

²⁷ <https://www.indicators.odyssee-mure.eu/energy-efficiency-database.html>. Accessed 29 March 2022

²⁸ <https://x-tendo.eu/about/>. Accessed 7 September 2021

²⁹ <https://cordis.europa.eu/project/id/845958>. Accessed 7 September 2021

a report in March 2020³⁰ assessing the status and potential of EPCs in Europe. According to the report, around 6 million residential EPCs are issued every year. Hurdles mentioned in reaching a comprehensive coverage of high-quality EPC data are:

- Inadequate data gathering
- Lack of granularity
- Lack of compliance
- Different EPC definitions and calculation methods

The report further investigates the quality assurance processes, calculation approaches, EPC databases, and EPC costs in the different countries and regions within countries. It also investigates public acceptance.

The X-tendo analysis highlights that EPCs are still commonly perceived as being unreliable, although some countries and regions have made efforts to overcome weaknesses. X-tendo recommends the following solutions to improve EPC reliability:

- Make EPCs more transparent and reliable.
- Tailor EPCs to the needs of the end user, displaying other indicators and conveying them in a user-friendly way.
- Use EPCs to deduce information on the building stock performance, potential of energy renovations, and indicators for which it is difficult to find data, such as Indoor Environmental Quality (IEQ).
- Make EPCs more dynamic by linking them with other resources, such as building logbooks and BRPs (Building Renovation Passports).
- Make EPCs accessible and useful to a broader client base, such as financiers, real estate agencies and contractors, by providing ready-to-use information in a database.
- Include new indicators such as real energy consumption, thereby promoting building smartness aspects such as demand response and dynamic pricing.
- Provide standards and guidance on how to calculate the EPC.

In addition to this list, BuiltHub recommends carefully investigating the link between EPCs and the real estate market. A survey conducted in the ZEBRA2020 project³¹ and published in a 2017 study³² comprising 618 interviews conducted with real estate agents from 8 EU Member States revealed that the main elements considered in selecting, purchasing, or leasing real estate were, in order of decreasing importance, *location* and *price*, followed by *size* and *nuisance* (busy road, landing airplanes, etc.). Energy performance was reported at the 10th position in the ranking. This finding also explains to some extent the reported result that buyers of a property disregard the energy performance rating in most cases.

³⁰ <https://www.bpie.eu/publication/energy-performance-certificates-in-europe-assessing-their-status-and-potential/>. Accessed 7 September 2021

³¹ <https://zebra2020.eu/>. Accessed 7 September 2021

³² Pascuas RP, Paoletti G, Lollini R. Impact and reliability of EPCs in the real estate market. Energy Procedia 140, 2017, 102-114. <https://doi.org/10.1016/j.egypro.2017.11.127>

According to the same study, even tenants renting a property do not pay much attention to the energy performance rating in most cases.

While increasing the reliability of EPCs may help in this regard, the impact of energy efficiency on the costs for the tenant should be investigated as well. If the energy-related share of the rent is low relative to other (fixed or variable) costs, most tenants would probably just ignore the energy performance rating.

Finally, the study reports that there is typically no increase in the rent for buildings with an excellent energy performance rating. While this is good news for tenants, owners may be less interested in energy-renovating their property because of this fact. This lack of incentive for owners is the so-called “tenant-landlord split incentive” or “dilemma”.

We note here that the recent energy crisis makes it necessary to follow up the above assessments with further studies.

Although the BSO contains certification as a thematic area, there is clearly a need to improve on the accessibility and completeness of the openly available data. A list of EPC registers and their level of accessibility is provided by the European Commission.³³ Still, a high number of countries provide no access to their EPC registers. This may be for several reasons ranging from privacy to technical issues. Some countries provide limited access or different kinds of accesses depending on the specific region and type of user. In Italy, for instance, authorities acting on behalf of municipalities, autonomous regions, and provinces can access the central EPC portal SIAPE³⁴ managed by ENEA (Italian National Agency for New Technologies, Energy and Sustainable Economic Development) to monitor the EPCs under their authority. They are also obliged to annually update the central portal with data from their cadastres and EPC registers. Citizens, companies, and other organisations do not have access to the data itself but can visualise aggregated statistics on the EPCs for some of the regions.

As of September 2021, nine countries reportedly offered a public EPC register, namely Denmark, Estonia, Ireland, Lithuania, the Netherlands, Portugal, Slovakia, Sweden, and the United Kingdom.³⁵ However, the level of accessibility and the procedures to access and reuse the data still varied considerably among these countries.

The following case studies on EPC registers refer to the status encountered in September 2021 considering an arbitrary external stakeholder without special prior information about the registers or access to them.

3.5.1. Case study #1: England and Wales

For England and Wales, upon registering on the “Energy Performance of Buildings Data” website hosted by the UK government³⁶ and agreeing to licencing and terms of use, it was

³³ https://ec.europa.eu/energy/content/public-epc-registers_en. Accessed 8 September 2021

³⁴ Sistema Informativo sugli Attestati di Prestazione Energetica, <https://siape.enea.it/>. Accessed 8 September 2021

³⁵ https://ec.europa.eu/energy/content/public-epc-registers_en. Accessed 20 October 2021

³⁶ <https://epc.opendatacommunities.org/>. Accessed 8 September 2021

possible to bulk-download an extensive registry of domestic single-building EPCs (about 4 GB of data). However, people can opt out of the public EPC register if they do not want other people to be able to see their EPC. In addition, non-domestic EPCs and Display Energy Certificates (DECs; records of the energy usage of public buildings) are available for download.

The downloaded domestic EPC data was organised in a series of folders, each referring to a single local authority and containing two files, one for the certificates themselves and one for improvement recommendations. Both kinds of files were given in CSV (Comma-Separated Values) format allowing for flexible and easy re-use. The provided information included, at single-building (more precisely, location by address) level:

- Detailed location (address, post code, etc.)
- Property type (mainly house or flat)
- Tenure
- Building type (mainly detached or semi-detached)
- Floor level and height
- Whether a flat is on the top storey
- Total floor area
- Construction age band
- Number of habitable and heated rooms
- Current and potential energy rating (from A to G)
- Current and potential energy efficiency and environmental impact (score from 1=worst to 100=best)
- Current and potential primary energy use (kWh/m² per year)
- Current and potential total CO₂ emissions in t/year and per unit floor area in kg/(m² year)
- Heating, hot water, and lighting current and potential annual cost
- Descriptions of technical systems (heating, hot water, lighting, secondary heating, main heating control) and construction elements (floor, walls, roof, windows) as enumeration of several categories
- Energy and environmental efficiency of technical systems (heating, hot water, lighting, secondary heating, main heating control) and construction elements (floor, walls, roof, windows) rated qualitatively from “very poor” to “very good”
- Type of glazing (single to triple) and estimated glazed area (qualitative, from “less than typical” to “much more than typical”)
- Electricity tariff type
- Access to the central natural gas network
- Main type of fuel used (enumeration of categories)
- Number of total and low-energy lighting outlets, percentage of low-energy lighting (derived from the number of outlets)
- Use of renewables (number of wind turbines, PV area as percentage of roof area, solar water heating)
- Whether a heat loss corridor is present along with corridor length
- Type of mechanical ventilation
- Inspection date
- Transaction type (e.g., rental, sale, new dwelling)

Recommendations given together with the certificates consist of up to 10 possible improvements for a single building. More than 60 different types of improvements are categorised, among which insulation, equipment or component substitution, installation of renewable energy systems, and smart control. For each improvement, an indicative cost range is supplied. However, this cost range is a rough estimate based on general types of renovation measures, i.e., it does not represent an estimate based on a specific calculation made for single buildings.

The England and Wales EPC register is one of the best EPC registers available. Nevertheless, there is room for improvement. Possibilities and recommendations for improving the EPC register in England and Wales (not the EPC data itself) from BuiltHub's viewpoint are to:

- Provide extensive metadata associated with the data files, following the guidelines given in BuiltHub Deliverable 3.1 "Inventory structure and main feature and datasets"³⁷.
- Provide ready and user-friendly access to material explaining the meaning of the indicators. User-friendly means that a clear explanation is given to non-expert users, while expert users find a transparent, detailed, and accurate explanation of the methodology including underlying assumptions and calculations.
- Provide analytics and visualisation tools.
- Allow programmatic access to data and metadata and exporting in common machine-readable formats that *include the semantics* (the meaning of the information exported) for automated correct interpretation and re-use of the data. This is commonly referred to as the **semantic web** and enabled through technologies such as RDF (Resource Description Framework) and OWL (Web Ontology Language).

To some extent, these recommendations are already implemented. For instance, after signing in on the EPC register, guidance is given on data coverage, data quality and the meaning of all data fields. Answers are given to frequent questions. For developers, the available APIs are described.

In addition to these functionalities offered by the central platform, some local authorities offer online data exploration in tables or graphs and expose an API documentation for programmatic access and software developers but finding and understanding the meaning of the data is not always straightforward. While the data type (numeric, categorical, etc.) and a basic description may commonly be found, detailed descriptions and semantics are often lacking.

3.5.2. Case study #2: The Netherlands

For the Netherlands, current energy labels could be downloaded online³⁸ both as "total" file with all currently valid labels and as daily "mutation" files containing changes to the "total" file. A new "total" file is published every first day of the month. To download the files, an API key was needed, which was sent by email after registration.

Data could be downloaded in CSV, XLSX, and XML format. Data fields contained:

³⁷ <https://builthub.eu/resource?uid=534>. Accessed 12 September 2023

³⁸ <https://www.ep-online.nl/PublicData>. Accessed 10 September 2021

- Building class (housing or commercial)
- Building type (apartment, detached, terraced, etc.)
- If apartment: position within building (top, middle, bottom floor; corner or in-between)
- Energy label (from G up to A++++)
- Usable floor area
- Energy demand
- Primary fossil energy
- Share of renewable energy
- Overheating temperature
- Heating demand
- Recording date

The website further offered a public REST API, also explaining some of the fields but not all and only in Dutch. Moreover, no units of measurement could be found, neither in the data files nor in the API documentation. The latter referred to the national standard, which must be purchased. The data files' field names were in Dutch only.

The BuiltHub recommendations are therefore to:

- Consider adding more data fields to the publicly downloadable certificate files, such as information about other final energy uses (e.g., hot water, lighting), technical systems, and building components' energy performance.
- Provide English labels and explanations for data fields as well, following recognised data schemas and ontologies.
- Provide units of measurement and other important information in the API itself so that platform users are not forced to buy national standards to be sure about the meaning of the data fields.

While the use of the Dutch public EPC register felt challenging to access and the data itself is difficult to reuse because of the above reasons, this is still one of the best examples of user-friendly open EPC registers in Europe currently accessible to the public.

3.5.3. Case study #3: Denmark

An online map³⁹ provided by the Danish Energy Agency was found showing energy labels for single buildings. Clicking on a label showed an image of the building, the building type (such as detached single-family house), the construction date, and the energy used for heating (e.g., natural gas). It is possible to go to a separate webpage with details for the specific building selected and the possibility to download the EPC itself in PDF format. Remarkably, all this was possible without registration.

In the downloaded EPC, the following additional information could be found:

- Energy label for heating (from A to G)
- Heated area

³⁹ <https://old.spareenergi.dk/demo/addresses/map>. Accessed 12 September 2023

- Annual heating expenditure including VAT and taxes
- Fuel consumption for heating
- Recommendations for improvement, indicating:
 - Type of measure (e.g., replacing the current gas boiler with a condensing one)
 - Expected energy and monetary savings
 - Estimated investment
 - Repayment time
 - Expected energy label after the improvement
- Status and suggestions for improvement (descriptive texts; e.g., “replace current double glazing with energy glazing”) for:
 - Building components (e.g., walls, windows)
 - Ventilation
 - Heating
 - Electricity
 - Water
 - Renewable Energy

BuiltHub’s first and foremost recommendations are to:

- Provide the numeric data in each EPC in a reusable data format (CSV, XML, or an RDF serialisation format) to allow analytics.
- Allow the download of multiple EPCs.
- Allow programmatic access to the EPC register (with machine-readable information).
- Offer an English translation of the EPC.

However, most recommendations given for the previous case studies apply to this case study as well.

3.5.4. Other countries

BuiltHub’s recommendations for the case studies reported in Sections 3.5.1-3.5.3 hold, adequately transposed, for the other European Member States. In summary, they advise to provide:

- Extensive, standardised metadata, see BuiltHub Deliverable 3.1 “Inventory structure and main feature and datasets”⁴⁰ for details.
- Full explanatory material in English on the meaning and measurement units of the data.
- Download of all available data in a common, reusable data format.
- Information on building geometry, constructive materials, technical systems, and energy performance.
- Analytics and visualisation tools.
- Programmatic access to data and metadata.

Among the countries stating that a public EPC register was available (see Section 3.5; reference date: September 2021), only aggregated data could be accessed in some cases, in

⁴⁰ <https://builthub.eu/resource?uid=534>. Accessed 12 September 2023

form of an image or through online interaction (e.g., by hovering over with the mouse) without the possibility to download the data points (the numerical values). For several countries, the EPC register could not readily be found via an Internet search. Language may constitute a barrier since all searches were formulated in English.

The recommendation in this regard is to provide an easy-to-find European central access point to EPC data and to allow programmatic access and download of numerical data for further re-use.

An effort in this direction was made by H2020 project Enerfund⁴¹ (end date: April 2019) in which they developed a tool to rate and score deep renovation opportunities based on EPCs among other parameters. Findings from this project were (Geissler, et al., 2019):

- Open access to EPC data “directly from the database of the competent authorities” was provided by Denmark, Bulgaria, the Netherlands, the United Kingdom.
- In Greece, France, and Romania, only aggregated results were publicly available.
- In the other countries, there was no public access to the EPC data.
- In none of the studied cases data was geocoded (allowed precise placement on a geographic map).
- “In some countries” (not better specified), EPCs were “not fully trusted” because of “energy performance calculation [...] done based on default values” and “the building [...] not even visited”.
- “EPC data [...] not harmonised across all Member States”, with “classification of each data category [...] not standardized”.

As recommendation, the authors concluded: “The analysis presented here shows that harmonization across Europe is urgently needed in two ways: first, regarding energy data as such, and second, regarding geo-coding and public accessibility”. They further recommended “to the European Commission and DG Energy in particular, that harmonisation of all energy data, not only EPC data, – and their geocoding – should become mandatory across all Member States.”

BuiltHub’s findings strengthen these recommendations and expand on them.

3.6. Data service providers

This section aims to map identified building-related data sources, service providers, platforms, and associated stakeholders. Starting from the needs and technical requirements of the stakeholders as given in D2.2, the aim is to show the potential of these resources to inform and provide knowledge about the European building stock.

The **EU Building Stock Observatory (BSO)** considerably supported this process by gathering data coming from different sources and projects, which is a huge undertaking. The great importance of the BSO also lies in the fact that the European Commission is committed to making it the central access point and ideally a one-stop-shop for European building stock

⁴¹ <https://cordis.europa.eu/project/id/695873>. Accessed 13 September 2023

data. In the September 2021 publicly accessible version of the BSO⁴², the data source for each type of data was typically indicated as a link to a project's website. However, the individual project's website might after a while not exist anymore or have been moved, or the datasets might no longer be accessible. For these reasons, it is oftentimes difficult if not impossible to re-use the data and even less so to understand the methodology behind it, such as how it has been collected or derived (e.g., through measurements, surveys, modelling) and how it should be interpreted.

Some paradigmatic examples of the issues encountered when searching for data sources underlying the BSO and possible basic short-term solutions are described below.

Total number of dwellings: the reported data source refers to the Odyssee⁴³ database except for Austria where a specific link is given. However, the link refers to an **Enerdata**⁴⁴ database, which is not accessible after registering on the Enerdata website. It is possible to ask for a seven-day free trial access to data about energy and emissions, energy modelling and forecasting, market analysis and intelligence, and energy efficiency and demand but afterwards a subscription is required. A sign up is also possible at the **Odyssee-Mure** website, offering free access for all EU Ministries, Concerted Action EED, EED Committee Members, EU universities and research centres for non-commercial uses, while subscription is required for other users. It is thus not straightforward to identify the specific author(s), understand more about the methodology used and assumptions made to calculate the data, know about the license and terms of re-use, or just find the same data in the Odyssee database. A basic solution would be to link to transparent metadata, documentation, and re-use information, such that the accessible numerical data is clearly understood and can safely be used for further processing, analytics, dissemination, etc.

Number of single-family dwellings (and other fields): the BSO reports as source "EC ENER/C3/2016-547/02/SI2.753931; EASME". The tooltip reveals the additional information "EU Commission: Comprehensive study of building energy renovation activities and the uptake of nearly zero-energy buildings in the EU; Hotmaps (CORDIS) 723677". With this information, those familiar with EU projects will probably know how to find the **Hotmaps** project website and database⁴⁵, but the process can be made more straightforward and easier for arbitrary users by linking to meta-information.

Number of detached houses (and other fields): the reported source is **Eurostat**, however, the link indicated on the BSO does not refer directly to explanations, metadata, or the original table(s) on Eurostat from which the data has been taken or derived. Therefore, this information cannot easily be found. Again, the basic solution is to provide detailed meta-information.

Number of detached houses (permanently occupied dwelling) (and other fields): for quite specific breakdowns, the BSO lets a user choose an indicator for which no data is available at

⁴² https://ec.europa.eu/energy/eu-buildings-database_en. Accessed 15 September 2021

⁴³ <https://www.odyssee-mure.eu/>. Accessed 14 September 2021

⁴⁴ <https://www.enerdata.net/>. Accessed 14 September 2021

⁴⁵ <https://www.hotmaps-project.eu/> and <https://www.hotmaps.eu/map>, respectively. Accessed 14 September 2021

all. These indicators should be hidden or marked in such a way that the user does not have to bother looking for non-available data.

Number of offices: this field is peculiar in that it is derived from different sources depending on the country, in this case from different national statistics sites. Also in this case, the link reported by the BSO is to the landing page of the organisation, with the advantage that the page is expected to remain accessible over time but making it difficult to identify explanatory information, metadata, and the underlying data itself. For a few countries, the data has been taken from the national LTRS. Finally, the numbers for several countries are reported to be estimates based on surface data and ratios from neighbouring countries. However, no reference is given as to how this calculation was performed, in which year, by whom, etc. As for the previous cases, detailed meta-information can resolve the issue.

Number of health care buildings: for most countries, the reported reference is the Hotmaps project but for The Netherlands there is a link to the open data on the Dutch statistics portal that can directly be compared with the data on the BSO. Furthermore, the Dutch portal gives a brief table explanation in English that includes references (mostly documents in Dutch) to the sources and methods to obtain the data. This can therefore be considered a good practise example.

Share of dwellings built before 1945: this data, as well as similarly detailed breakdowns, typically comes from national statistics databases. Therefore, almost all countries have their own reference. The quality of the references varies somewhat from country to country. Typically, it is the landing page of the national statistics institute where in the best case the database can be easily found and data, metadata, and documentation retrieved. For some countries, the user is referred to the Odyssee project or a PDF document. This plethora of references allows for a transparent view and gives interesting information about available data platforms. Such a register of references/links to databases is by itself a very useful resource. However, an additional central explanation given on the BSO website is recommended to make the observatory more attractive and user-friendly. Also, a direct link with the **European Census Hub**⁴⁶ (or the Eurostat database) seems desirable to ensure data coherence.

Share of dwellings in densely-populated area (urban centre): the given link leads to another table in the Eurostat database, namely to the “Number of private households by household composition, number of children and age of youngest child (1 000)” (data code: lfst_hhnhtych).

Share of owner-occupied dwellings: the source is reported as JRC/EASME. However, the links and tooltips mention the 2011 census data, the Hotmaps project, and Eurostat. While this is probably just a bug to be fixed, the point that can be made here is that even if several data sources are brought together to compute an indicator, it is once more important to give a reference to meta-information, such as metadata about authors and publication date along with documentation on the methodology used.

Share of tenant occupied dwellings (rent at market price): data is available for the year 2017. The unit of measurement is indicated as “%”, however data is above 100 in most cases and

⁴⁶ <https://ec.europa.eu/CensusHub2/query.do?step=selectHyperCube&qhc=false>. Accessed 14 September 2021

ranges widely between around 20 and more than 25000. It is guessed that a mismatch between data and metadata or data and indicator happened. The recommendation given is to add automatic data checks to the manual ones whenever possible. In this specific case, the share indicated as percentage should fall in the range 0-100.

Share of tenant occupied dwellings (rent at reduced price or free): the correct link to the Eurostat table is given. This allows a verification of the value, which is important. Unfortunately, a random spot verification reveals that the values reported for Austria on the BSO correspond to the values in the Eurostat table for the tenure status “Owner” (and not “tenant” under specific renting conditions). Such errors are difficult to track, the suggestion is therefore to interconnect databases (in this case, the BSO and Eurostat database) as much as possible, i.e., to keep only one “master” repository (possibly with mirrors to reduce the risk of data loss and inaccessibility) and let “slave” repositories access the data from the master repository, with well-defined and well-documented data transformation processes from the master to the slave repository if needed.

Share of dwellings with single-person households: Eurostat is reported as source, but any meta-information is missing (see recommendations given above).

As of September 2023, the BSO is accessible from a different website⁴⁷ and is being reworked in a tender⁴⁸ lasting until 2026, with new data and a new interface.

In addition to the data sources referred to by the BSO, several others have been identified in BuiltHub, see D3.1 Annex A.

3.7. Existing European data strategies and roadmaps

This section explores data strategies and roadmaps the BuiltHub roadmap draws on. In many cases, buildings represent a specialisation and application of general data strategies and roadmaps. Existing efforts are investigated towards a sustainable building data value chain, to understand the barriers, roadblocks, and challenges these efforts have encountered so far and what solutions they propose to overcome them. The latter, combined with the BuiltHub results, constitute the ground truth for the BuiltHub recommended measures and roadmap for a self-sustained building data value chain, see Section 4.

A milestone for implementing a big data framework in Europe was the FP7 BIG project (GA no. 318062)⁴⁹ lasting from 2012-2014. The project had the objective to provide a **roadmap for usage and exploitation of big data**. The efforts were published in a book (Cavanillas, Curry, & Wahlster, 2016) showing the vast potential associable with big data and specifically focusing on:

- Opportunities

⁴⁷ https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/eu-building-stock-observatory_en. Accessed 13 September 2023

⁴⁸ Tender CINEA/2022/OP/0010 Continuation and Improvement of the EU Building Stock Observatory (LIFE 2021-2024)

⁴⁹ <https://cordis.europa.eu/project/id/318062>, <https://www.big-project.eu/index.html>. Accessed 26 October 2021

- The big data value chain, from acquisition to usage
- Usage and exploitation in different sectors (industrial, health, public, finance and insurance, energy and transport, media and entertainment)
- Roadmaps and action plans for technology, businesses, policy, and society

Table 1 shows that data management processes, quality checks, security and privacy, visualisation and user experience, and analytics are strongly required across sectors, an outcome that is also confirmed by the results obtained in BuiltHub emerging from deliverable D2.2 “Report on relationships of stakeholder needs and requirements”.

Table 1: Sectorial requirements taken from (Cavanillas, Curry, & Wahlster, 2016), p. 265

Technological Requirement	Number of Demanding Sectors	Sectors						
		Health	Public	Finance & Insurance	Energy & Transport	Telecom & Media	Retail	Manufacturing
Data Management Engineering	3		X			X	X	
Data Enrichment	2	X				X		
Data Integration	5	X	X	X		X	X	
Data Sharing	4	X	X	X	X			
Real-Time Data Transmission	3		X				X	X
Data Quality	3	X		X			X	
Data Improvement	2					X	X	
Data Security and Privacy	7	X	X	X	X	X	X	X
Data Visualization and User Experience	2						X	X
Deep Data Analytics	3		X	X			X	
Modelling Simulation	3		X				X	X
Natural Language Analytics	3		X				X	X
Pattern Discovery	3		X	X		X		
Predictive Analytics	2		X			X		
Prescriptive Analytics	3				X	X	X	
Real-Time Insights	5		X		X	X	X	X
Usage Analytics	2			X		X		

Further, the book gives a prioritisation of the requirements shown in Table 1, see (Cavanillas, Curry, & Wahlster, 2016), Table 15.2, p. 274. The highest priority level “urgent” is assigned to

data security and privacy, integration, real-time insights through analytics, and data sharing. Data quality is assigned the priority level “important”.

Technology, business, policy, and society roadmaps and associated actions are given for a timeframe from 2015 (one year after the end of the FP7 project BIG) to 2020, see (Cavanillas, Curry, & Wahlster, 2016), pp. 279-286, and brought together in an integrated roadmap, which for clarity is reproduced in Figure 2.

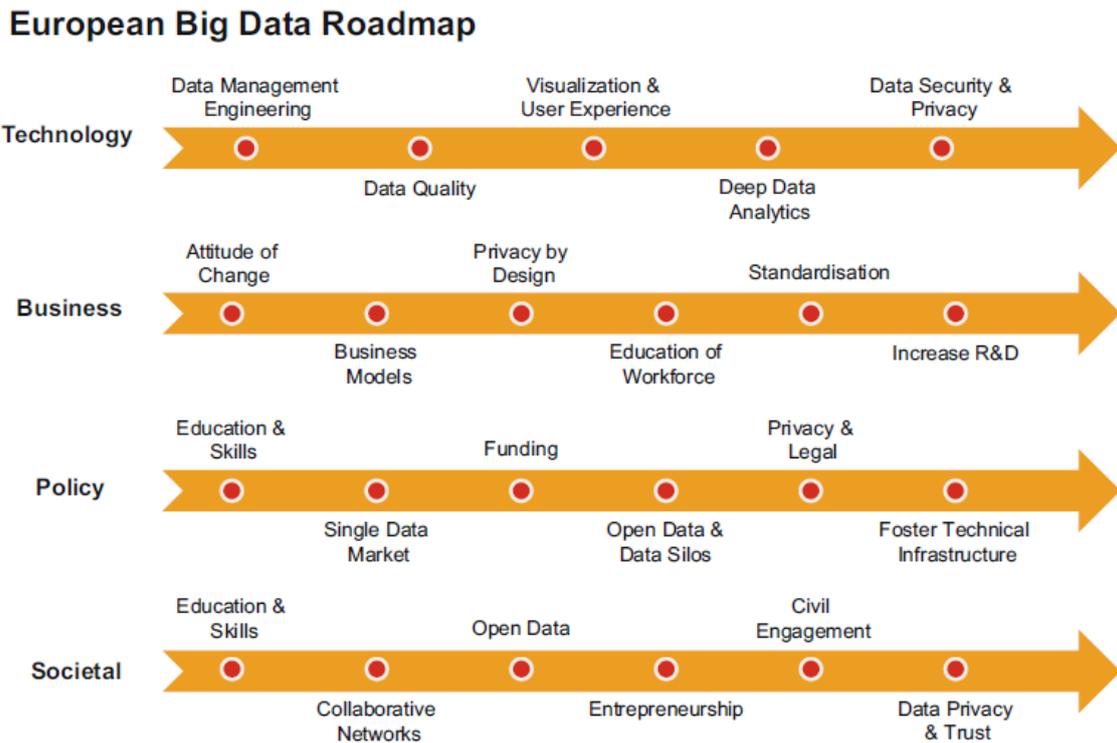


Figure 2: Big data roadmap for Europe taken from (Cavanillas, Curry, & Wahlster, 2016), p. 286

Another H2020 project that developed solutions to overcome barriers related to exploiting big data analytics was the TOREADOR project⁵⁰, which pointed out that companies not specialised in the Big Data Analytics (BDA) sector “may find it difficult to evaluate whether the advantage they can obtain from tapping into Big Data sources is commensurate to BDA costs”⁵¹. Further, the project investigated how legal concerns about data sharing can be dispelled.

While many roadblocks and challenges remain, several steps have been taken towards these goals in the last years. The European data strategy⁵² is a noteworthy step towards recognising the importance of data and data sharing for Europe. A public consultation has been carried out

⁵⁰ <http://www.toreador-project.eu/> and <https://cordis.europa.eu/project/id/688797>. Accessed 3 November 2021

⁵¹ <http://www.toreador-project.eu/objectives/>. Accessed 3 November 2021

⁵² <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy>. Accessed 20 October 2021

in 2020⁵³ with the aim to create a common European data market in which data can be safely exchanged according to the FAIR (Findable, Accessible, Interoperable, Reusable) data principle. This consultation led to a proposal for a Regulation on European data governance⁵⁴, which aims to address public sector data re-use, business-to-business data sharing, personal data usage in compliance with GDPR, and data use on altruistic grounds.

The European data strategy rests on four pillars:

1. Governance framework for data access and use
2. Enablers, i.e., investments in data infrastructure, technology, and capabilities
3. Data-related competences and skills of individuals and SMEs
4. Data spaces for nine sectors, among which the domains green deal, mobility, and energy, promoting data sharing and usage between sectors

Several associations and initiatives aim to accelerate the implementation of a data strategy, among which the LODIC (Linked Open Data Interest Community)⁵⁵, IBPDI (International Building Performance & Data Initiative)⁵⁶, IDSA (International Data Spaces Association)⁵⁷, GAIA-X⁵⁸, Data Sharing Coalition⁵⁹, SWIPO⁶⁰, BDVA (Big Data Value Association)⁶¹, AIOTI (Alliance for Internet of Things Innovation)⁶², and FIWARE⁶³. This typically happens through a combination of actions like the publication of position papers, the creation of networks, hubs, or communities bringing together stakeholders willing to contribute, the establishment of working groups with specialized knowledge in different domains, the exchange of experts in events, and the development of technical solutions. For instance, the International Data Spaces Association published several position papers, white papers, and technical documents on a Reference Architecture Model (IDS RAM), the connection between the IDS RAM and the GAIA-X architecture, the required provisions and agreements for data sovereignty and exchange, the design principles for data spaces, and use cases. While the latter comprise many different sectors not necessarily closely related to buildings, some of them are, such as mobility, air quality including CO₂ measurement, and smart cities.⁶⁴ In September 2021, BDVA, FIWARE- GAIA-X, and IDSA joined forces launching the Data Spaces Business Alliance (DSBA).⁶⁵ IBPDI with its 50+ members, among which leading international industry, has

⁵³ https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12491-Data-sharing-in-the-EU-common-European-data-spaces-new-rules_en. Accessed 20 October 2021

⁵⁴ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767>. Accessed 20 October 2021

⁵⁵ https://ec.europa.eu/eurostat/cros/content/linked-open-data-interest-community_en. Accessed 28 October 2021

⁵⁶ <https://ibpdi.org/>. Accessed 4 November 2021

⁵⁷ <https://internationaldataspaces.org/>. Accessed 25 October 2021

⁵⁸ <https://www.data-infrastructure.eu/GAIA-X/Navigation/EN/Home/home.html>. Accessed 25 October 2021

⁵⁹ <https://datasharingcoalition.eu/>. Accessed 25 October 2021

⁶⁰ <https://swipo.eu/>. Accessed 26 October 2021

⁶¹ <https://www.bdva.eu/>. Accessed 26 October 2021

⁶² <https://aioti.eu/>. Accessed 29 October 2021

⁶³ <https://www.fiware.org/>. Accessed 26 October 2021

⁶⁴ International Data Spaces Association, Data Spaces Brochure 2021, Available at <https://internationaldataspaces.org/publications/most-important-documents/>. Accessed 26 October 2021

⁶⁵ <https://internationaldataspaces.org/bdva-fiware-gaia-x-and-idsa-launch-alliance-to-accelerate-business-transformation-in-the-data-economy/>. Accessed 26 October 2021

created a strong network working on a global data standard for the real estate and has published an openly available Common Data Model (CDM)⁶⁶. This CDM puts the concept of digital building twin at the centre, see Figure 3, and builds on Microsoft’s CDM as part of Microsoft’s Open Data Campaign that was launched in 2020⁶⁷. Related to digital twin technology, the Digital Twin Hub⁶⁸ has pointed out that an initiative parallel to that of the IBPDI exists, namely the Real Estate Core ontology⁶⁹ drawing on existing W3C ontologies for buildings and building systems.

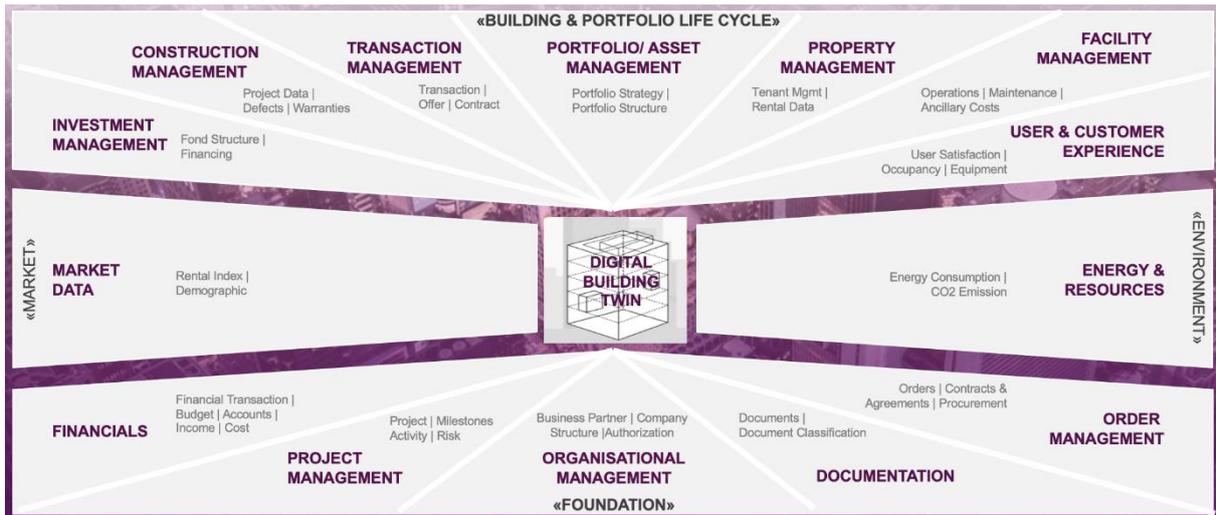


Figure 3: Data clusters in the Common Data Model for real estate developed by the IBPDI⁷⁰

Another key pillar in standardising the work with data, especially geospatial data, is given by the INSPIRE⁷¹ (Infrastructure for Spatial Information in Europe) Directive. This spatial data infrastructure helps to put into effect cross-border environmental policies or actions and is based on the corresponding infrastructures by the Member States. The Directive addresses 34 spatial data themes needed for environmental applications including buildings. Buildings are among the themes of INSPIRE, which is therefore an excellent resource for a standardised building data specification. It is thus important to follow this specification or suggest improvements to develop and promote a harmonised approach at EU level.

Due to its nature, the INSPIRE data specification on buildings establishes a strong link with geographic and geometric data, thereby embedding a building in a spatial context and spatially representing building data. To illustrate this point, we briefly present the use case on “Energy / Sustainable Buildings” given in Annex B.4.4 of the “D2.8.III.2 INSPIRE Data Specification on Buildings – Technical Guidelines”⁷² (European Commission Joint Research Centre, 2013).

⁶⁶ <https://github.com/ibpdi/cdm>. Accessed 4 November 2021

⁶⁷ <https://blogs.microsoft.com/on-the-issues/2020/04/21/open-data-campaign-divide/>. Accessed 5 November 2021

⁶⁸ <https://www.digitaltwinhub.org/>. Accessed 5 November 2021

⁶⁹ <https://digitaltwinhub.co.uk/a-survey-of-idms-and-rdls-dt-build/>. Accessed 5 November 2021

⁷⁰ Image taken from: <https://github.com/ibpdi/cdm/blob/main/README.md>. Accessed 10 November 2021

⁷¹ <https://inspire.ec.europa.eu/>. Accessed 1 October 2021

⁷² <https://inspire.ec.europa.eu/id/document/tq/bu>. Accessed 27 October 2021

Starting from a UML use case diagram showing the main stakeholders involved in building energy consumption and sustainability assessment and promotion (Figure 4), the following steps are reported:

- Retrieve the associated INSPIRE themes, in this case BU (buildings) and US (utility and governmental services)
- Identify the buildings to process, using building attributes such as 2-d geometry, official area, and current use of the building
- Decide on the building performance and sustainability evaluation and calculation methods
- Collect input data required to perform the evaluation
- Update INSPIRE data with the result of the evaluation

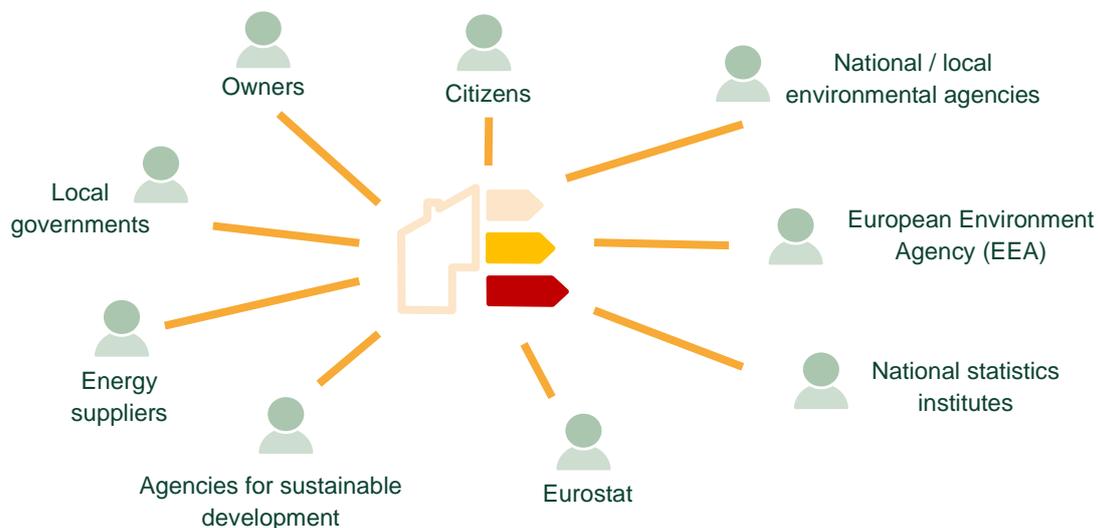


Figure 4: Stakeholders of building energy performance and sustainability assessment and promotion. Adapted from (European Commission Joint Research Centre, 2013).

Further, it is stated that some input and output data can be naturally matched with readily available INSPIRE data, such as 2-d building geometry, official area, current use, service type, number of floors, date of construction, external reference to the official building register or permit, and energy performance. However, it is also stated that some information may not be found within the INSPIRE framework, thus requiring other repositories or completion by field surveys.

From the above considerations it becomes clear once more that an effort is needed to bring together building data from multiple sources. This is reiterated by the Linked Building Data (LBD) Community Group hosted by the World Wide Web Consortium (W3C)⁷³. The group consists of experts in BIM (Building Information Modelling) and the so-called *Web of Data*. A key idea of the latter is to move away from traditional tabular databases and data embedded in documents and spreadsheets and to shift to graph data with nodes and labelled directed links between nodes. This allows for a more flexible structure able to integrate heterogeneous

⁷³ <https://www.w3.org/community/lbd/>. Accessed 27 October 2021

data and a potentially faster processing. The concepts of a *semantic web*, web of linked data, Web of Things (WoT), and Internet of Things (IoT) are closely related since the nodes are interpreted as data (as abstraction of a physical object receiving, storing, processing, and sending data) and the links as descriptors of relationships between data, thereby giving the data a meaning interpretable by machines. Outputs of the LBD Community Group are a Git repository⁷⁴ with mainly ontologies and tools and engagement actions with stakeholders through surveys and events jointly with the LDAC (Linked Data in Architecture and Construction) workshop series⁷⁵. In these events, real-life applications and industry cases are presented and discussed as well as research topics.

Further, there is a vast array of efforts to improve data-driven services at single building to building cluster level, among which the three EU “Big data for buildings” projects BEYOND, BIGG, and MATRYCS⁷⁶ focus on leveraging and tapping into the large potential of big data. Use and business cases are experimented and demonstrated in several large-scale demonstrators across Europe and across market actors and include:

- Benchmarking and building portfolio management
- Assets management (contracts, invoices, equipment, meters, sensors, etc.)
- Certification, considering not only energy but also other indicators, such as the sustainability indicators of Level(s)⁷⁷
- Building performance monitoring
- Energy conservation measures tracking and evaluation
- Lifecycle thinking, enabled by interoperability between data hubs, physical systems, and tools for building design, operation and management, and investment
- Energy performance contract management
- Occupant-centric building design
- Building operation optimisation through weather, occupancy, and energy price forecasts
- Demand-response and demand-side flexibility

Cross-cutting topics for all use cases are digitalisation, standardisation, security, privacy, interoperability, and data ownership protection, items which are dealt with in the three projects.

Another set of projects focuses on next-generation EPC, namely, iBRoad2EPC, crossCert, TIMEPAC, EuB SuperHub, CHRONICLE, SmartLivingEPC, D2EPC, EDYCE, ePANACEA, EPC RECAST, QualDeEPC, U-CERT, and X-tendo⁷⁸. A transversal theme in these projects is making EPCs more dynamic and closer to real building performance, and increasing their

⁷⁴ <https://github.com/orgs/w3c-lbd-cg/repositories>. Accessed 27 October 2021

⁷⁵ <http://www.linkedbuildingdata.net/ldac/index.html>. Accessed 27 October 2021

⁷⁶ BEYOND: <https://cordis.europa.eu/project/id/957020>. BIGG: <https://cordis.europa.eu/project/id/957047>. MATRYCS: <https://cordis.europa.eu/project/id/101000158>. Accessed 28 October 2021

⁷⁷ https://ec.europa.eu/environment/levels_en. Accessed 28 October 2021

⁷⁸ iBRoad2EPC: <https://ibroad2epc.eu>. crossCert: <https://www.crosscert.eu>. TIMEPAC: <https://timepac.eu>. EuB SuperHub: <https://eubsuperhub.eu/>. CHRONICLE: <https://www.chronicle-project.eu/>. SmartLivingEPC: <https://www.smartlivingepc.eu/en>. Accessed 14 September 2023. D2EPC: <https://www.d2epc.eu/en>. EDYCE: <https://edyce.eu/>. ePANACEA: <https://epanacea.eu/>. EPC RECAST: <https://epc-recast.eu/>. QualDeEPC: <https://qualdeepc.eu/>. U-CERT: <https://u-certproject.eu/>. X-tendo: <https://x-tendo.eu/>. Accessed 29 October 2021

transparency, reliability, trustworthiness, and ultimately value and relevance in the market. This is done by:

- Standardising the methodology – input data to be collected, monitoring and calculation methods, reporting, etc. – for a convergence of EPC schemes at EU level.
- Leveraging building data collection infrastructure and management systems.
- Working with a digital twin (i.e., a virtual representation close to reality) of the building.
- Reporting building performance in terms of a set of energy, environmental, financial, IEQ / comfort / well-being, smartness (e.g., through the SRI), and flexibility KPIs.
- Streamlining the issuance and periodic update of EPCs.
- Transparent and friendly engagement with the end-user.
- Inspection or auditing procedures.
- Linking the EPC with other property data stores, such as passports, logbooks, notebooks, renovation roadmaps, and the like; see, for example, the project iBRoad⁷⁹ devoted to the implementation of building renovation passports and building logbooks, and its successor iBRoad2EPC merging EPCs with the building renovation passport and logbook.

In summary, while information about the EU building stock is fragmented, incomplete, and still needs to live up to its full potential, best practices are being demonstrated or proposed in multiple initiatives and projects that show how at least some of the issues can be overcome and how opportunities could be exploited.

Challenges encountered can be grouped into different categories. From the start to the end of the data value chain, in chronological order:

- Data identification – which data should be collected and what for?
- Data collection
- Data processing
- Data analysis
- Data monetisation

The above-mentioned projects provided BuiltHub with insights for all these steps. Since BuiltHub's mission was not to collect new data but to define a roadmap for durable data flow, BuiltHub did not exhaustively apply all the methodologies for increasing the value of the data indicated in the other projects to the data sources mentioned by them. Rather, BuiltHub performed a selection of promising data sources and strategies confirmed in stakeholder engagement actions carried out in WP2 and experimented this selection on the BuiltHub web-based platform developed in WP5. Data assembly and analysis strategies are reported in the deliverables of WP3 and WP4, respectively.

Section 4 presents the BuiltHub solutions for a self-sustained building data value chain that came out from this process.

⁷⁹ <https://ibroad-project.eu/>. Accessed 1 October 2021

4. Recommended measures for a self-sustained building data value chain

A successful open data approach needs to be rooted in a functioning data value chain, that is, the value of data for the stakeholders should steadily increase from collection to generation of impact. Furthermore, for the value chain to be self-sustained there must be valid benefits for all stakeholders involved. BuiltHub confidential deliverable D6.1 “BuiltHub platform value proposition” investigates benefits for lead and end users generated from BuiltHub’s results, e.g., from the offered data services enabling knowledge development. Also, it is deemed important to raise the stakeholders’ interests in FAIRification of data by acting upon multiple (economic, environmental, social, etc.) drivers. Every stakeholder should identify one or more tangible benefits for themselves. For individuals, this can be an improvement of their daily work or private lives. For a public or private organisation, there may be economic or other benefits for their business.

For the analysis of the driving factors involved, in BuiltHub we adopt a PESTLE (Political, Economic, Social, Technological, Legal, Environmental) approach and detail measures for each of these dimensions. The business models are treated for the most part as economic dimension. With this approach, BuiltHub is largely in line with the European Big Data Roadmap in Figure 2, which shows the PEST dimensions if we map “Business” with “Economic” and “Societal” with “Social”. While the legal aspect can be seen as incorporated in the “Policy” dimension, we can see in Figure 2 that the environmental aspect is missing in the European Big Data Roadmap or at least hidden behind other factors in the spotlight, although it is frequently mentioned as primary driver in EU policy today and may remain a primary driver in the long term.

As can be seen in the following subsections, the dimensions are interwoven and affect and support each other.

4.1. Political

Member States and their public authorities can benefit from a **mandate** from the EU that allows them to pursue efforts in data collection and management such that they can legitimately dedicate resources to this important endeavour of data collection and processing. Moreover, they need **guidelines** or **directives** on how to transform their data infrastructure and processes, what data to collect and under what terms, and how data can/should be made available for further use.

It is important to strike the right balance between dictating strict rules and offering flexibility to Member States. Too much flexibility in handling and sharing data or knowledge can hamper interoperability between data services and comparability of data analysis. On the other hand, too many or too strict rules may not accommodate enough the differences between Member States and create disadvantages for some of them. Excessive rules may also smother innovation in, e.g., data analysis and reporting.

A mandate combined with lean directives and clear guidelines could overcome these hurdles. However, when given a mandate, Member States should be made aware of and fully

understand the advantages and correct ways of following such a mandate. Otherwise, the required data collection and management efforts would be perceived as a rather pointless duty and would therefore be executed poorly. Once the Member States have clear in mind the advantages and a possible transformation plan, it is important that dialogue with stakeholders ranging from companies to the citizen is sought, to meet local societal requirements and needs and to achieve social acceptance. **Active stakeholder participation** and several feedback rounds are recommended at each stage of the process, from inception to operation.

Further, dialogues with the BuiltHub community have revealed that there is a strong need for an **overarching standard on data management for the entire building lifecycle** as well as guidance on what standards should be followed. Standards for specific categories certainly exist, see e.g.:

- ISO/TC 59/SC 13⁸⁰ on “Organization and digitization of information on buildings and civil engineering, including BIM”
- ISO/TC 184⁸¹ on “Automation systems and integration”
- CEN/TC442 on “Building Information Modelling (BIM)”
- Industry Foundation Classes (IFC)⁸² and Green Building XML (gbXML) Schema⁸³ for interoperability between building design and engineering software
- International Cost Management Standard (ICMS) for benchmarking and reporting of construction project cost⁸⁴

However, it is difficult for smaller companies to understand what standards are strategic to pursue, to assure compliance with those standards, and to develop products compatible with competing standards. It is therefore important to support efforts towards the creation of one encompassing standard (or a family of complementary, non-overlapping standards) and to produce **guidelines and tools** for the correct use of such a standard or family of standards. Regarding tools, a notable effort is that of the EPB Center⁸⁵, which offers spreadsheet tools for the demonstration of several standards.

A proper shaping and execution of building data processes requires instructed personnel. Therefore, another important part of the political dimension is **training the workforce** in the building sector in the creation, management, use, and exploitation of building data assets. These trainings should be accompanied by appropriate **education pathways** for forming professionals and awareness raising of the public in the medium to long term.

The development of **technical infrastructure** in terms of storage, computational power, transaction speed, security, stability, service-oriented architecture, and user-friendly software needs to accommodate the expansion of the data value chain. Policies should allow this

⁸⁰ <https://www.iso.org/committee/49180.html>. Accessed 10 November 2021

⁸¹ <https://www.iso.org/committee/54110.html>. Accessed 10 November 2021

⁸² <https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes/>. Accessed 10 November 2021

⁸³ <https://www.gbxml.org/>. Accessed 10 November 2021

⁸⁴ <https://icms-coalition.org/>. Accessed 10 November 2021

⁸⁵ <https://epb.center/>. Accessed 10 November 2021

development to take place within a structured framework under EU sovereignty and foster efforts that are coherent with and accessible to societal needs.

4.2. Economic

Promising **business models** must underpin sustained data collection, data sharing, and related services. Data collection, curation, storage, sharing, as well as creating added value out of data and ultimately providing useful services to stakeholders requires lots of resources and is time intensive. Therefore, **enablers** are investigated, evaluated, and proposed that allow economic viability in the long-term, e.g.:

- *Obligation*
 - How it works: a mandate to collect and manage data is given, typically accompanied by appropriate financial aid coupled with a compliance-checking mechanism.
 - Example: the European Commission asks Member States to collect, manage, and report data about their building stocks.
- *Data exchange model*
 - How it works: companies create data marketplaces or exchanges where data providers can share their data with other businesses in exchange for access to other datasets, services, or compensation.
 - Example: a consortium of property management companies creates a data exchange platform where they share building performance data, such as energy usage and maintenance records, with each other. This enables benchmarking and best practices sharing among participating companies. Similarly, governments of Member States could share data about their building stocks with each other, for the same reasons and benefits.
- *Freemium*
 - How it works: companies offer a basic service for free, but users can upgrade to a premium version with additional features or functionality by paying or sharing their data.
 - Example: BuiltHub is making a free basic version of its platform public access to that allows users to monitor energy usage and temperature. In exchange for sharing more detailed building data, such as occupancy patterns or HVAC system performance, building owners can upgrade to a premium version with advanced analytics and predictive maintenance features.
- *Data monetisation*: companies collect user data and then sell or license that data to third parties, such as advertisers, researchers, or marketers.
- *Subscription with data insights*: companies offer subscription-based services and provide data insights or analytics to subscribers based on their usage of the service.
- *Personalization and recommendation*: companies use customer data to personalize their products or services, making recommendations or customizing content based on user preferences.
- *Data-for-discount*: companies offer discounts, coupons, or rewards to customers in exchange for sharing their data or participating in data-gathering activities.

- *Crowdsourced data*: companies collect data from a large group of users who voluntarily contribute data in exchange for access to a collective dataset or a service.
- *Data collaboration*
 - How it works: companies collaborate with each other to share data that can benefit both parties, often for research, market analysis, or business insights.
 - Example: administrations provide cadastral and EPC data to energy agencies and research institutions in exchange for the development of energy renovation scenarios.
- *Data-as-a-Service (DaaS)*
 - How it works: companies offer data-related services, such as data hosting, cleansing, enrichment, quality checking, visualisation, or API access, in exchange for a fee.
 - Example: the BuiltHub consortium merged different datasets together to enrich building stock data with additional indicators derived from calculations on several merged datasets. The BuiltHub platform allows hosting, filtering, visualisation, and download of the data.

The BuiltHub's point of view is that a business model will only work if each public stakeholder involved is granted legitimacy, capability, and means to carry out the required activities while clearly seeing the benefits. Private organisations need to see clear benefits for their business as well, economic, or otherwise. Finally, individuals must perceive participating in the data value chain as beneficial for their work or private life.

The BuiltHub business models to be developed as part of Task 6.3 are closely linked to the BuiltHub community needs and requirements, platform services, use cases, and pilot initiatives. Deliverable D3.4 "Functionalities of the BuiltHub platform and provided services" (currently only available on the EC Participant Portal) reports the provided services and business approach for each stakeholder type based on the BuiltHub community inputs given in deliverable D2.2 "Report on relationships of stakeholder needs and requirements". The following organisation types are part of the BuiltHub community:

- Academia
- Financial
- Designer
- Policy maker
- Energy agency
- Public body
- Industry
- Consultancy
- Non-profit / association

Furthermore, BuiltHub is constantly reaching out to address other types of stakeholders. The above organisations are also consulted to provide insight on stakeholders in their network that are difficult to engage directly.

In the following, we give an example of a possible service concept and business approach emerging from the stakeholder engagements. Others are contemplated in above mentioned BuiltHub deliverable D3.4 “Functionalities of the BuiltHub platform and provided services”.

We consider a real estate developer or owner of a building portfolio interested in energy retrofitting their properties to improve indoor comfort, save on energy bills, and raise the value of the properties. The retrofit process may typically include the following steps:

- Analysis of pre-retrofit building status
 - Analysis of official documentation (cadastral data, technical documentation of installed components, etc.)
 - On-site inspection
 - Status report
- Development of upgrade plan proposal
 - Proposed improvements
 - Expected energy savings
 - Financial analysis (e.g., capex, opex, ROI)
- Retrofit design
 - Design project
 - Tender documentation
- Tender, bid, and procurement management
- Installation, commissioning
- Delivery
- Support

The analysis of a building’s status and history can be supported by a building logbook acting as repository of building plans, previous works, energy consumption and production profiles, etc.

BuiltHub recommends linking an energy retrofit project to certification and rating processes, as the retrofit should lead to tangible improvements in energy performance and other aspects reflected by the certificate and rating. The dimensions analysed and recommendations given in this deliverable facilitate dynamic certification, rating, and retrofit processes, which today are burdened by cumbersome procedures in data management. This means that the property owner must pay much more for the service or invest much more time in preparing documents or both. From the owner’s perspective, transitioning from paper-based, manual, non-standardised, opaque procedures to a streamlined, transparent, trustworthy workflow is advantageous if there is a gain in time or money or both. An important benefit might also be the lower risk of issues arising during the process. From the service provider’s perspective, at first sight there might seem to be no incentive to lower the price of the service. However, assuming a sufficiently flexible market, a service provider can benefit from offering owners lower prices in exchange of getting more and better-structured data. The service can be offered at a lower price thanks to a streamlined data management, which allows service providers to gain a competitive edge, i.e., to outperform competitors. Furthermore, the data from owners can be sold or used to further improve the offered service.

4.3. Social

BuiltHub's proposed roadmap can only succeed if its principles and practical steps are accepted by society. A **shift in the mindset towards a data-sharing culture** driven by mutual benefits is required. Such a culture must become second nature to people.

This cultural and social transition should be supported by other dimensions as follows.

- **Political**
 - A training framework for professionals and proper education of the public is needed to allow building of competences and skills related to working with building and building-related data.
 - Supporting the creation of an open, accessible, and flexible data infrastructure and market will allow people to flexibly work with data.
- **Economic** have proven a crucial motivation for people to provide and share data that otherwise is difficult or costly to obtain. People are willing to share even quite sensitive data if the provided service derived from the data is perceived as worth it. Thus, a key motivation can also be other non-monetary benefits, for instance, homeowners or shop owners may be willing to publicly expose the energy performance, IEQ, or other KPIs of their homes or shops, respectively, if this unlocks or attracts services that are interesting for them.
- **Technological** improvements in digitalisation, accessibility, security by design (as enabled, e.g., by blockchain technology), privacy by design, and user-friendliness are required to gain the trust of people that their data is adequately and carefully handled according to agreement.
- The **legal** framework protects and defends data rights building trust.
- Potential **environmental** implications of improved monitoring and planning of the building stock transition on daily life must be tracked and communicated in a transparent manner during awareness-raising, training, and education.

Humans spend most of their time in buildings.⁸⁶ With the increasing number of smart metering and sensing devices installed in buildings and available to occupants, buildings can be operated more efficiently, and occupants can be made aware of energy consumption, IEQ, and costs. This allows them to make choices beneficial to their lives and the environment. Therefore, BuiltHub supports the rollout of smart infrastructure (e.g., smart metering), systems, and devices. Although efforts exist to make IEQ data available to all, see for instance the ASHRAE Global Thermal Comfort Database II⁸⁷, the collection of IEQ data is still limited and not included in the BSO. If the BuiltHub recommendations are followed, data from buildings and platforms are expected to become FAIRer (more findable, accessible, interoperable, and reusable). This data could then be used for example to provide benchmarks on IEQ to property owners but also citizens in general.

⁸⁶ Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, Behar JV, Hern SC, Engelmann WH. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *J Expo Anal Environ Epidemiol.* 2001 May-Jun;11(3):231-52. doi: 10.1038/sj.jea.7500165. PMID: 11477521.

⁸⁷ <http://www.comfortdatabase.com/>. Accessed 10 November 2021

Benchmarks make sense for many other indicators referenced in deliverable D3.1, such as energy consumption per square meter in total and by end use (space heating and cooling, DHW, lighting, etc.), building shell performance (U-values for roof, walls, floors, windows), and CO₂ emissions per square meter.

Public authorities, policy makers, building energy auditors and certifiers, manufacturers, and researchers can utilise these benchmarks to assess the evolution of the building stock over time, design and revise policies, identify risks and improvement possibilities, and assess and enhance products and services to be offered on the market.

The transformation of the building value chain will entail a transformation of jobs dealing with building data. Menial jobs are expected to be less needed and to shift towards higher-qualified tasks, which on the one hand will require proper education and training but on the other hand will offer higher job security and salaries.

4.4. Technological

A first step along the technological path is **digitalisation**, to enable fast and easy processing and re-use of data. Manual and paper-based procedures should generally be avoided, because they are labour-intensive, slow, error-prone, and render efficient data processing and re-use very difficult if not impossible.

As an example, a potential for improvement is given by many of today's building rating and performance certification processes which, according to input from the BuiltHub community and specifically real estate managers and building owners, require a high amount of paperwork. This increases the time and cost of the process itself and hinders re-use of information.

Only in specific cases, for example when dealing with highly sensitive information where illicit tracking, hacking, or tracing would have severe consequences, paper documents may be considered a better choice. They might also serve as hardcopy for contracts, although the advent of digital signatures, blockchain, and other cryptosecurity solutions are making this purpose more and more obsolete.

However, digitalisation alone is not enough. The traditional “web of documents” allows for processing of information by humans but not necessarily by machines (computers and other devices). Many of the highly used document formats today, such as DOC, TXT, PDF, and HTML, do not allow machine-readability, i.e., easy extraction of structured data and information by machines. Therefore, a step towards encoding of semantics (meaning) with the data (the “semantic web”) or inferring semantics from the data (recent ML and, generally, AI approaches) is needed. The “semantic web”, an expression coined in 2001⁸⁸, is an extension of the World Wide Web standardised by the W3C. Technologies such as RDF and OWL are central elements since they can formally represent metadata. Together with other structures

⁸⁸ Berners-Lee, Tim; James Hendler; Ora Lassila (May 17, 2001). "The Semantic Web". Scientific American. Retrieved 2 November, 2021.

(taxonomies, ontologies, rules, etc.), these technologies aim to facilitate the integration of information from different sources.

Since more recently, ML and AI offer a modern take on the idea of the semantic web while responding to some of the criticisms regarding the semantic web's implementation complexity and other issues. Although the most adequate guidelines and technological solutions are under constant development, BuiltHub sees a clear necessity and trend towards **interoperable, standardised machine-driven data communication and services**.

Today, data acquisition in buildings is facilitated by the increasingly common use of monitoring through meters, sensors, BMS, smart equipment and devices, wearables, etc. However, the data collected is oftentimes not exposed to an interested stakeholder but buried in proprietary systems for system-internal use. Even if the data is accessible, this does not mean that it is exploited in the best way. For instance, with the rollout of smart electricity meters, there is the potential to better inform electricity suppliers about consumer behaviour, which would allow them in turn to offer the consumer more attractive billing from profit generated by optimising the central system. This potential remains largely untapped if the demand profiles of consumers are not constantly analysed.

The success of the BSO crucially depends on convincingly delivering the technical requirements stated in Table 1. From a functional architecture point of view, the BuiltHub web-based platform consists of six areas to deliver these requirements:

- Data gathering
- Data pre-processing and quality checks
- Data ingestion
- Data storage
- Service execution
- Service availability

The data gathering can occur either programmatically, through an API in an automated or semi-automated fashion by machine-to-machine communication, or manually, by a user uploading data to the BuiltHub platform. This *raw* data must then go through the data pre-processing and **quality checks** (see deliverable D3.2 "Methodology on quality assurance"⁸⁹) to become suitable for ingestion into the database of the platform. The service execution and service availability areas ensure that services such as data analytics and visualisation are available to all stakeholders anytime and that they are promptly executed. While the data gathering can be considered as one end of the platform exposed to the outside world, the service availability area represents the other end where a machine or a human can interact with the platform. BuiltHub recommends this functional architecture for the next-generation BSO, as it is expected to accomplish the technical requirements in Table 1.

For the BuiltHub platform, a graph database was chosen at the heart of the architecture to deal with the huge variety of data involved and since it allows semantic queries.

Furthermore, technical solutions that allow overcoming barriers related to privacy, data ownership protection, and IP protection should also be investigated (and deployed once mature), as well as those, that allow stakeholders to share data without having to renounce the value of their data.

4.5. Legal

Data providers, users, service providers, and intermediaries need transparent legislation on data provision, sharing, and re-use agreements and contracting schemes and procedures, tackling aspects such as **privacy, ownership protection, intellectual property protection, and licensing**. Private companies in the data business and private owners of data need **clear rules** on these aspects.

To allow rapid and easy integration of these schemes and procedures in their daily business, guidance should be offered to them in multiple ways, for instance through document templates, consultancy, training, best practice examples, and online forms, preferably offered by online or physical one-stop shops.

BuiltHub proposes a generic **data provision model agreement** and **IP rights strategies** for adaptation by stakeholders.

The political dimension should ensure the allocation of appropriate funding and resources to the creation and provision of document templates and training facilities, monitoring and impartial evaluation of existing efforts, and compilation of best practices.

4.6. Environmental

In BuiltHub, the primary driver for wanting to revolutionise the building value chain is the belief that such a revolution will have far-reaching benefits at all levels of society. Massive **energy savings and decarbonisation** are expected to be triggered by leveraging building-related data, because this data will allow greener building design, construction, operation, renovation, and end of life. Moreover, better decisions will be taken at energy infrastructure and community level.

Decarbonisation is a key element in mitigating climate change and needs to be reconciled with the rapid growth of global energy demand. The European Environment Agency (EEA) maintains a database on climate change mitigation policies and measures in Europe⁹⁰ to reduce greenhouse gas (GHG) emissions. The reported strategies include:

- Retrofitting buildings to make them more energy efficient
- Adopting renewable energy sources
- Developing more sustainable transport
- Promoting sustainable land use

⁹⁰ <http://pam.apps.eea.europa.eu/>. Accessed 10 November 2021

Buildings are a central element of all items in the above bullet list because they are prosumers of renewable energy sources, represent a central factor in e-mobility, and compete for land use.

The political, economic, societal, technological, and legal measures recommended by BuiltHub are devised to drive decarbonisation and consequently generate a strong positive impact on the environment. This impact is expected to largely offset the negative environmental impact caused by the additionally required digital data infrastructure. Not to be neglected is the environmental benefit of reducing paper and physical traffic linked with paperwork.

5. Implementation and conclusion

5.1. Roadmap timeline in brief

In this section, the recommended measures of Section 4 are embedded in a **timeline for implementation** shown in Figure 5.

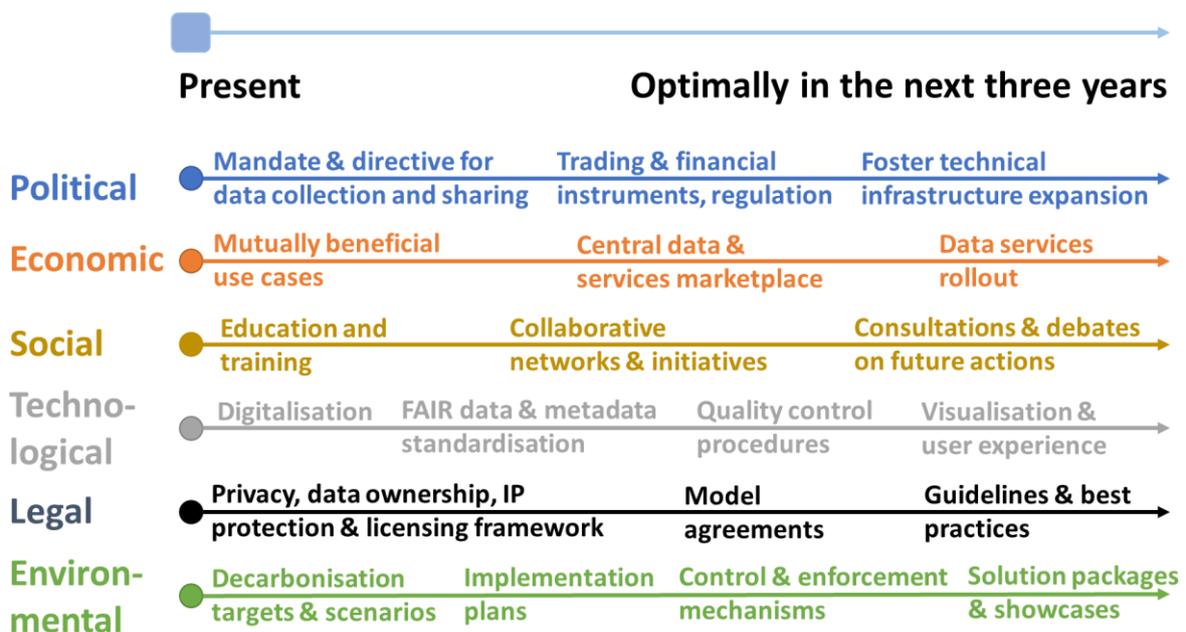


Figure 5: Roadmap timeline

This timeline gives a recommendation on the **sequence of actions** to be taken for the different dimensions (from political to environmental). For each dimension, each “thread” of actions (i.e., each horizontal line in Figure 5) may be developed almost independently from the other threads, although interactions can prove to be useful. For each thread, earlier actions prepare the terrain for later actions and thus act as prerequisites. Also, while the threads are represented as straight horizontal lines, they can be considered as loops undergoing multiple iterations and revisions. For instance:

- Political – mandates & directives are the basis for establishing supporting and regulating instruments operating in line with the former. Such instruments will allow

controlled (regulated) improvements in building data collection and sharing. Mandates or directives for data collection and sharing are expected to foster *harmonised* gathering and storing of information in the different Member States, which will facilitate the aggregation process at EU level. For instance, defining which indicators to report, and at which level of detail, gives Member States more clarity on what kind of data they should collect and what information should be generated from the data. An example of such an effort is the Concerted Action EPBD (CA EPBD) Key Indicators and Decisions (KI&Ds)⁹¹ presenting relevant EPBD performance indicators or implementation decisions per country. These indicators and decisions can change over time to incorporate new policies, advancements in technologies and construction practices, etc.

- Social – consultations and debates on future actions are going to shape education and training. For example, the focus on specific technologies in a decarbonisation path would lead to increased education and training in the deployment and use of those technologies.

The timeframe considers an optimal case, which would see a first implementation wave of the reported measures within three years although it is deemed highly probable from the consortium's experience that the process will take considerably longer in practice. The timeframe of three years has been chosen for the following main reasons:

- To achieve the decarbonisation targets envisaged by the EC, the fastest possible action is needed.
- There is no clearly insurmountable a-priori obstacle to performing these measures in the next three years.

This timeline is largely in line with the roadmap presented in (Cavanillas, Curry, & Wahlster, 2016) but updates and expands on it.

Realistically, the implementation process of the measures in Figure 5 is expected to take longer than three years considering that in (Cavanillas, Curry, & Wahlster, 2016) the year 2019 after a four-year implementation starting in 2015 was chosen as milestone year for extremely ambitious targets envisioning Europe as world leader in several big data domains.

The following sections expand on Figure 5 with more detailed information on recommended next steps.

5.2. Recommended steps to undertake

Figure 6 shows in a simplified manner main steps to undertake for deploying a data services platform such as the BuiltHub web-based platform and the BSO.

⁹¹ <https://epbd-ca.eu/database-of-outputs>. Accessed 4 November 2022

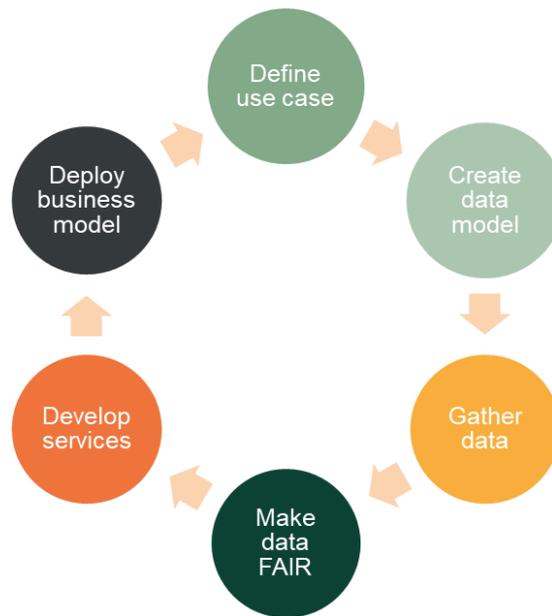


Figure 6: Main steps in the data value chain

The **use case** is the starting point. This includes identifying the stakeholders, with their needs and requirements, and the goals to achieve. For example, a use case is the creation of a LTRS (or its evolution, the National Building Renovation Plan, NBRP, as proposed for the revised EPBD⁹²). Stakeholders involved in gathering and processing the required data are typically the government, energy agencies, housing institutes, business institutes, engineering institutes, academia, etc. The goals may be to characterise:

- The building stock itself. This may entail identifying features to which specific types of retrofit packages can be applied, and clustering buildings into groups according to those features.
- The most suitable retrofit packages to apply to specific types of buildings (buildings belonging to a previously identified cluster) in terms of, e.g., energy savings, avoided GHG emissions, and costs.
- The policies and measures needed to deploy the retrofit packages.

When building the use case, *awareness raising* may be necessary, as stakeholders may not know what goals can be achieved and how. Thus, potential opportunities and benefits should be explored and discussed together with them.

Next, to allow integration into existing data and knowledge exchange frameworks and platforms, the envisioned use case and goals must be cast into a well-defined **data model**. Here, data model is intended in a broad sense including intermediate data and transformations in the data modelling process. Creating the data model involves several steps:

⁹² Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the energy performance of buildings (recast). COM/2021/802 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0802>. Accessed 8 November 2022

- Definition of the sought-after output indicators. For example, a renovation plan may consider as outputs for the different scenarios final energy consumption or savings, CO₂ emissions, and investment costs, broken down by sector, end use, and energy source. Furthermore, retrofits may be characterised by the expected post-retrofit performances of building components (e.g., U-values of roof, façade, or slab post-retrofit insulation, new windows) and equipment (e.g., COP of a heat pump replacing a gas boiler). For each output indicator, an unambiguous definition and documentation is required. Indicators and definitions should be taken as much as possible from existing authoritative sources and standards. For a common EU BSO to succeed, priority should be given to directives and standards at EU level or endorsed by the EU.
- Producing an output indicator will typically require many intermediate steps, input data, and calculations. For the purposes of this section, any data transformation and all generated intermediate indicators from primary raw data to the output indicator are part of the data model. Therefore, all transformations and intermediate indicators need to be treated in the same rigorous fashion as the output indicators. Specifically, they must be defined unambiguously, be well-documented, and integrated into existing terminology and ontologies as much as possible.

To co-create a data model and prepare the following data gathering with stakeholders requires establishing a process in which a common understanding and language (i.e., a shared terminology) is established about the data to collect and how to collect it. This process may start with the creation of working groups for the different sectors and consist of several iterations in which the working groups propose indicators and transformations and experiment, discuss, and revise them.

With the data model defined, the **data gathering** process can be streamlined to follow the data model, see Section 5.2.1. In the best scenario, gathered data is directly suitable for ingestion, i.e., to be transferred to its destination (such as a database, data warehouse, data mart, or document store).

If the public is involved in the data gathering (e.g., citizens filling out forms), proper awareness raising should be carried out and clear instructions should be given beforehand to ensure that they can adequately fulfil their task.

If use case and data model have been defined properly, **making data FAIR** is greatly facilitated, especially as regards interoperability and reusability, see Section 5.2.2. However, data and indicators will also be *found* more easily if they follow existing terminology and vocabularies. Accessibility requires some more steps, such as stable and secure infrastructure and services, authorisation mechanisms, and publishing data in compliance with GDPR, IPR protection, and licenses.

Access to data alone is a value by itself, especially if the data is combined from different sources. However, processing and analysing data, including quality control procedures, most often requires a lot of expertise and hence technical sectoral experts and data specialists such as data analysts and data scientists. Therefore, the **development of data services** is the next important step. Today, entities involved in building stock data analysis such as statistics institutes and energy agencies may have inhouse data specialists or rely on external services.

The BuiltHub project contributes to the development and demonstration of services with several deliverables, in particular:

- D3.2 “Methodology on quality assurance” (public)
- D3.4 “Functionalities of the BuiltHub platform and provided services” (public)
- D4.4 “Description of a workflow how to implement and transform data” (confidential)
- D4.5 “Report on data visualization” (public)
- D5.1-D5.4 on the BuiltHub platform (confidential)

The last step in the simplified data value chain diagram shown in Figure 6 is the deployment of a business model allowing sustained provision of data services. In this step, ways to raise the necessary budget to cover expenses and allow for investments for the provided data services must be studied and implemented, see Section 5.2.3.

If the business model proves to be unsuccessful, previous steps in the data value chain must be revisited, possibly starting again from the use case.

5.2.1. Gathering of primary data

Arguably, the biggest issue still faced today is the lack of primary data. This issue was already pointed out by the EPISCOPE project in 2016 in their 4th Synthesis Report on “Tracking of Energy Performance Indicators in Residential Building Stocks - Different Approaches and Common Results”, pp. 93-100.⁹³ They concluded that:

- “Available information sources are not sufficient to fulfil the prominent role they should play for climate protection strategies [...]. There are wide information gaps concerning the actual state as well as the trends of building thermal insulation and efficient / renewable heating systems.”
- “[...] there are many studies using building stock energy balance models [...]. But information gaps are often closed by estimations, which are not based on objective knowledge. [...] However, estimations [...] cannot replace the necessary objective observations. The basic problem [...] is the lack of reliable and comprehensive primary data.”

From BuiltHub’s analysis of the current BSO, this situation is basically unchanged. The EPISCOPE report proceeds by mapping and analysing **applicable data sources**, reporting that:

- Complete **inventory counts and building portfolios** were available only for two countries (CZ, FR). There were still relevant data gaps, flawed data, and insufficiently

⁹³ https://episcope.eu/fileadmin/episcope/public/docs/reports/EPISCOPE_SR4_Monitoring.pdf.

organised data maintenance. EPISCOPE concluded that “(almost) complete covering of the building stock alone is not sufficient. The data also need to be maintained”.

- Use of **censuses** is not a realistic option in most cases because of the small number and types of indicators considered and the difficulty to adapt to aspects such as energy efficiency and climate protection.
- **EPC databases** “might be an interesting option”. However, “EPC data might be not comprehensive” and include only calculation results but not underlying input data. Furthermore, EPC data might give a biased view of the building stock, providing incorrect statistical information. Correction factors may be used to reduce this risk but require validation through additional data sources. Representative surveys were suggested for that purpose.
- **Sample surveys** “play a very relevant role”, especially if a full portfolio is unavailable, which is “typically the case in large building stocks”. It was remarked the importance of following the principles of sampling theory and that “under these circumstances unbiased results can be expected and [...] standard errors [...] estimated”. However: “In most of the cases available data of the surveys was incomplete [...], not statistically robust [...] or out of date.” Overall, “a [...] comprehensive and broadly based survey approach can be strongly recommended”; and it “is very recommendable to include an as large as possible number of the interesting properties in the same survey so that also correlations can be analysed. Merging the data from different sources is always a difficulty, usually not making possible the observation of interrelations of the different variables.” The adequacy of even small sampling fractions was stated, with the caveat that less frequent building stock characteristics require a larger sample.

BuiltHub seconds these findings. Summarising:

- Potentially existing national inventories (such as EPC databases, building renovation passports, digital logbooks, cadastral data) covering a large fraction of the building stock must be quality-controlled and regularly updated, e.g., in case of building retrofits. They must be equipped with a mechanism for keeping track of updates (e.g., a version control system).
- Sample surveys, if done respecting statistics principles, are a reliable and effective (also economically) method to characterise the building stock.

GIS data is worth mentioning as further data source, which is increasingly gaining importance also thanks to services such as Google Maps and OpenStreetMap.⁹⁴ This data can be combined with **satellite/aerial and street imagery**. However, the creation of reliable 3-d representations of buildings and the extraction of constructive features represents a challenge.

In the future, **BIM data** may become a useful source, especially if it contains the relevant parameters, such as U-values.

⁹⁴ <https://www.google.com/maps> and <https://www.openstreetmap.org/>, respectively. Accessed 4 November 2022

Section 2.5 mentioned main actors who should be appointed – through appropriate directives, regulations, tenders, and other similar mechanisms – or incentivised – through financial instruments or other kinds of benefits – to gather primary data and be responsible for data quality and maintenance. Of utmost importance is that the data gathering supports data FAIRness, see Section 5.2.2.

5.2.2. Making gathered data FAIR

As mentioned in Section 3.7, FAIR data refers to **Findability, Accessibility, Interoperability, and Reusability** of data.

For interoperability and reusability, primary data must be gathered in a *structured* and *harmonised* way. For the gathering of primary data, see Section 5.2.1.

Deliverables D3.1 “Inventory structure and main feature and datasets” and D4.1 “Building sector indicators, definition, calculation, representation” list the indicators to report, and how they are calculated and represented. Countries therefore need to establish primary data gathering and pre-processing methods allowing such reporting. If there is agreement on the indicators, indicator calculation methods, and indicator representations among all countries, aggregation at EU level is greatly facilitated.

The confidential deliverable D4.3 “Data model design, specifications, mappings and storage requirements” tackles the problem of creating a standardised ontology for interoperability of data platforms. Main outcomes can be summarised as follows:

- ISO and EU specifications differ in many aspects. In conflicting cases, BuiltHub has given priority to EU rules and the latest DCAT-AP (Data Catalog Vocabulary Application Profile) and GeoDCAT-AP specifications.⁹⁵
- EU vocabularies and ontologies were followed, namely, Eurostat’s own vocabularies, such as SIEC (Standard international energy product classification)⁹⁶, NUTS (Nomenclature of Territorial Units for Statistics), and Publications (for countries and places)⁹⁷.
- A new vocabulary “blthb” was created for BuiltHub where existing vocabularies were not specific enough for the datasets integrated into the BuiltHub web-based platform. The vocabulary is used to represent concepts regarding energy, building materials, building types, building sectors, etc.
- The Data Catalogue present on the BuiltHub web-based platform specifies for each BuiltHub dataset its structure, indicating for each field title, general description, predicate, data type, cardinality, and an example for its content. Fields common to all

⁹⁵ <http://data.europa.eu/r5r> and <http://data.europa.eu/930>, respectively. Accessed 7 November 2022

⁹⁶ <http://dd.eionet.europa.eu/vocabulary/eurostat/siec/>. Accessed 7 November 2022

⁹⁷ <http://publications.europa.eu/resource/authority/country/>. Accessed 7 November 2022

datasets are spatial, temporal, and unit of measurement predicates. The data types or ranges are specified as well.

- To control **access and availability**, several *entities* – the representation of a concept, metric or key point found inside a dataset from a business/functional point of view – were defined. Building type, building sector, dataset type, access type, or availability type are examples of entities since specific businesses might get access to certain entities but not others.

5.2.3. Sustainably providing data services

Section 2.3 has stated several basic and more advanced services building on top of existing data. Such services, and especially more advanced services, can only be successful if there is a sound and high-quality base of primary data, see Sections 5.2.1 and 5.2.2.

A basic and crucial service that should not be forgotten is keeping data processes and platforms alive and well-maintained.

A tiered pricing model is recommended, in which basic services are offered to a large user base for free or at a very low price, more advanced services are offered to users willing to pay more, and highly resource-demanding, very advanced services are offered to customers able to sustain the price. Moreover, data can serve as currency insofar as data providers may get access to some data services “for free”, i.e., without any other form of payment in addition to the provided data.

BuiltHub confidential deliverable D6.1 “BuiltHub platform value proposition” presents three basic **value proposition models** for the BuiltHub web-based platform:

- 1) **Enabler**. The offered service consists mainly in enabling users to do more advanced data manipulations with their datasets. In other words, customers perform the main value-adding work on their datasets themselves.
- 2) **Expert system**. The offered service is mainly to answer users’ requests for specific results to be obtained by manipulating their datasets. That is, the customer provides the dataset and requests a result; the work is done by experts on behalf of the customer.
- 3) **Marketplace**. The offered service is configured as matchmaking system where the data provider is brought together with the data analyst or scientist.

Mixed value proposition models are possible to a certain extent. In BuiltHub, services provided by the BuiltHub web-based platform are presented in deliverable D3.4 “Functionalities of the BuiltHub platform and provided services”. Advanced services are described in the confidential deliverable D4.4 “Description of a workflow how to implement and transform data”. The latter document explores **ML models** in detail. Specifically, it:

- Maps use cases in building stock research to suitable ML methods.

- Gives practical examples drawing from available data in Sweden on the recycling potential of PVC flooring and hazardous material assessment in renovation and demolition.
- Develops a ML approach as part of the BuiltHub roadmap (this deliverable) and applies it to a pilot case, namely, the Swedish building database for energy retrofitting. This pilot case aims to enrich the Swedish EPC database and hence represents an important replication potential for other countries, provided that sufficient sound-quality primary data is gathered by those countries (see Section 5.2.1). The pilot case also demonstrates how street imagery can be leveraged to identify building characteristics an EPC is typically missing but are required to assign the appropriate energy retrofit package to the building.

BuiltHub explores (and recommends) automating parts of the ML workflow to bring down costs. However, a considerable amount of expert knowledge is typically required to carry out ML studies at high quality. Therefore, a two-level strategy is pursued:

- 1) Basic automated ML embedded in the data platform, allowing users to autonomously run customizable studies. There is an increasing amount of free or easily affordable ML libraries for common programming languages available on the market. Such a service can be offered at a more affordable price.
- 2) (Ideally building on Step 1) Commissioned, more complex ML analysis performed by experts. This service will typically have to be rather expensive to cover all costs.

Costs for the provision of data services can be contained further by leveraging, supporting, and participating in well-established frameworks for collaboration, infrastructures, processes, and tools. Frameworks at EU level are listed in Section 2.1. For instance, in the Concerted Actions, periodic reviews and discussions take place on implementation decisions and key indicators concerning building stock monitoring and transition. Regarding infrastructure, we mention the “official portal for European data”⁹⁸ as effort to provide persistent access to datasets including documentation and publications, stories, use cases, courses, and standards. Finally, the BuiltHub platform documented in the confidential deliverables D5.1-D5.4 represents what BuiltHub considers as state-of-art in deployable tools.

5.3. Conclusion

While the importance of a building data landscape is recognised and has shown promising developments in recent years, Europe still has a long and winding road ahead until full and mature development of a self-sustained building data value chain. In this deliverable we proposed measures (Section 4) and implementation steps (Section 5) towards this goal starting from the state-of-art and considering actual developments and emerging trends. If the European Commission and the EU Member States follow these implementation steps, BuiltHub expects an important positive impact on the European building stock transition.

⁹⁸ <https://data.europa.eu/en>. Accessed 15 November 2022

6. References

- Cavanillas, J., Curry, E., & Wahlster, W. (2016). *New Horizons for a Data-Driven Economy. A Roadmap for Usage and Exploitation of Big Data in Europe*. Switzerland: Springer.
- European Commission Joint Research Centre. (2013). D2.8.III.2 INSPIRE Data Specification on Buildings – Technical Guidelines.
- Geissler, S., Androutsopoulos, A., Charalambides, A. G., Jareño Escudero, C., Jensen, O. M., Kyriacou, O., & Petran, H. (2019). ENERFUND - Identifying and rating deep renovation opportunities. *IOP Conf. Ser.: Earth Environ. Sci.*, 323, 012174. doi:10.1088/1755-1315/323/1/012174
- R2M Solution, VITO, BPIE. (2020). *Study on the Development of a European Union Framework for Digital Building Logbooks. Final Report*. Brussels: European Commission.

